

GenCore version 5.1.4.P5-4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 9, 2003, 11:39:55 ; Search time 32.9328 Seconds

(without alignments)
3666.364 Million cell updates/sec

File: US-10-092-390-4

Perfect score: 3601
Sequence: 1 MVISLNSCLSFICLLCHWMT.....HCDSYCAEGRMGPRNCISLPCY 586

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SPREMBL_21:
1: sp_archaea:
2: sp_bacteria:
3: sp_fungi:
4: sp_human:
5: sp_invertebrate:
6: sp_mammal:
7: sp_mhc:
8: sp_organelle:
9: sp_phage:
10: sp_plant:
11: sp_protist:
12: sp_virus:
13: sp_vertebrate:
14: sp_unclassified:
15: sp_virus:
16: sp_bacteriophage:
17: sp_archaeap:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3601	100.0	1140	4	Q96K67
2	3468	96.3	567	4	Q8WUJ3
3	2094	58.2	969	4	Q96K66
4	1828	50.8	747	11	Q8VHE4
5	1824	50.8	1034	11	Q8VHL7
6	1824	50.7	1034	11	Q8VHL7
7	1372.5	38.1	1574	11	Q88281
8	1340	37.2	1664	5	Q9TVQ2
9	1309	36.4	220	11	Q63404
10	1284.5	35.7	1045	5	Q8T3A6
11	1284.5	35.7	1070	5	Q8T3A7
12	1284.5	35.7	1111	5	Q9XMD6
13	1282.5	35.6	1246	4	Q75095
14	1252	34.8	434	5	Q9W0A0
15	741	20.6	594	5	Q9W0A1
16	740	20.5	594	5	Q9Y151

17	708.5	19.7	2447	13	013149	013149 fuqu rubrip
18	685.5	19.0	2524	5	Q9GPA5	Q9GPA5 brachiolesto
19	682	18.9	4288	4	Q9NFK9	Q9NFK9 homo sapien
20	681.5	18.9	2531	5	016004	016004 lytechinus
21	681	18.9	2653	5	Q25253	Q25253 lucilia cup
22	678	18.8	1193	13	Q90819	Q90819 gallus gall
23	676.5	18.8	4006	11	Q35452	Q35452 mus musculu
24	672.5	18.7	4135	6	018977	018977 bos taurus
25	667.5	18.5	4114	11	054796	054796 mus musculu
26	664	18.4	1214	13	Q9Y0D2	Q9Y0D2 xenopus lae
27	648.5	18.0	1254	13	Q9YH02	Q9YH02 brachydanio
28	648.5	18.0	1254	13	Q90Y56	Q90Y56 brachydanio
29	648.5	18.0	2352	5	061240	061240 halocynthia
30	646.5	18.0	713	5	Q96289	Q96289 podocoryne
31	646	17.9	1532	13	Q90984	Q90984 gallus gall
32	646	17.9	1714	13	Q90995	Q90995 gallus gall
33	641.5	17.8	2019	11	064706	064706 mus musculu
34	638	17.7	3695	4	Q8TDF8	Q8TDF8 homo sapien
35	636	17.7	1216	13	Q90824	Q90824 gallus gall
36	628.5	17.5	1216	13	Q90Y55	Q90Y55 brachydanio
37	622.5	17.3	2146	5	Q9VC97	Q9VC97 drosophila
38	621	17.2	752	13	042374	042374 brachydanio
39	609	16.9	3319	5	Q9VJ75	Q9VJ75 drosophila
40	609	16.9	3367	5	Q9XZC9	Q9XZC9 drosophila
41	595	16.5	1212	13	042347	042347 gallus gall
42	593.5	16.5	3704	5	P91904	P91904 caenorhabdi
43	586	16.3	1722	5	Q19350	Q19350 caenorhabdi
44	583.5	16.2	2656	5	Q9GN03	Q9GN03 paracentrot
45	581	16.1	1404	5	Q9VB65	Q9VB65 drosophila

ALIGNMENTS

RESULT 1	
Q96K67	PRELIMINARY; PRT; 1140 AA.
AC Q96K67:	
DT 01-DEC-2001 (TREMBLrel. 19, Created)	
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)	
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)	
DE MEGF10 protein (K1A1780).	
GN MEGF10.	
OS Homo sapiens (Human).	
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
OX NCBI_TaxID=9606;	
RN [1]	
RP SEQUENCE FROM N.A.	
RC TISSUE=HIPPOCAMPUS;	
RX MEDLINE=24245430; PubMed=11347906;	
FA Nagase T., Nakayama M., Nakajima D., Kikuno K., Ohara O.;	
RT Prediction of the coding sequences of unidentified human genes. XX.	
RT The complete sequences of 100 new cDNA clones from brain which code	
RT for large proteins in vitro.	
DR EMBL: AB058676; BABF7409.1;	
DR InterPro: IPR000561; EGF-like.	
DR Pfam: PF00008; EGF_15.	
DR PROSITE: PS00022; EGF_1; UNKNOWN_17.	
DR PROSITE: PS01186; EGF_2; UNKNOWN_17.	
SQ SEQUENCE 1140 AA; 122204 MW; 4552FA239423895A CRC64;	
Query Match	100.0%; Score 3601; DB 4; Length 1140;
Best Local Similarity	100.0%; Pred. No. 0;
Matches 586; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY 1 MVISLNSCLSFICLLCHWITASPVLNEDNVCSHMSYSVTYQESYPRFDDIYITSC 60	
DB 1 MVISLNSCLSFICLLCHWITASPVLNEDNVCSHMSYSVTYQESYPRFDDIYITSC 60	
OY 61 TDILNMFCTGHRVSYRTAVRHGKTMVRRSQCPCPYEGSGMCVPCACVHGRCIA 120	

```

Db 61 TDILMWFCTRRHVSRYTRAYRHGEKTMRYRKSQCCPGFYESEEMCVPHCADRCVHGRCTA 120
Oy 121 PNTCCPEPMGGTNCSSACDGDHMGPHCTSRCCCKNGALCNPTTGACHCAAGFRGMRCD 180
Db 121 PNTCCPEPMGGTNCSSACDGDHMGPHCTSRCCCKNGALCNPTTGACHCAAGFRGMRCD 180
Oy 181 RCEOGTYGNDCHQRCQCCNGATCDHYTGECRCRPPGYTGAFCEDLCPGKHGPOCEQRCP 240
Db 181 RCEOGTYGNDCHQRCQCCNGATCDHYTGECRCRPPGYTGAFCEDLCPGKHGPOCEQRCP 240
Oy 241 QNGGVCHHTTGBCSCPSGSMGTVCQPCPEGRGRKNCSEBCCCHNGTCDATGQCHCSP 300
Db 241 QNGGVCHHTTGBCSCPSGSMGTVCQPCPEGRGRKNCSEBCCCHNGTCDATGQCHCSP 300
Oy 301 GTTGECODECPVGTGYVCAETCCQVNGKCYHYSGALCEAGFAGRCERLCEPGLY 360
Db 301 GTTGECODECPVGTGYVCAETCCQVNGKCYHYSGALCEAGFAGRCERLCEPGLY 360
Oy 361 GIKCDKRCPCHEHTHSHCHPMSEGCACRPGMSGLYCNETCSPGFYGEACQOITSCQNGAD 420
Db 361 GIKCDKRCPCHEHTHSHCHPMSEGCACRPGMSGLYCNETCSPGFYGEACQOITSCQNGAD 420
Oy 421 CDSYTKCTCAGPFGKIDSTPCPLGTGYGINSRCCKNDVAVCSYVDSCTCKAGMHGV 480
Db 421 CDSYTKCTCAGPFGKIDSTPCPLGTGYGINSRCCKNDVAVCSYVDSCTCKAGMHGV 480
Oy 481 DCSIRPSTGTWFGCNLTQCCLNGACNTLDGTCTCAPMRGEKCELPQODTYGGLNCAE 540
Db 481 DCSIRPSTGTWFGCNLTQCCLNGACNTLDGTCTCAPMRGEKCELPQODTYGGLNCAE 540
Oy 541 RDCSHADCGHPTTGHCRLPGMSGYHCDSCVCAEGRWGNCISLPCY 566
Db 541 RDCSHADCGHPTTGHCRLPGMSGYHCDSCVCAEGRWGNCISLPCY 566

RESULT 2
O8WUL3 PRELIMINARY: PRT: 567 AA.
ID O8WUL3
AC O8WUL3:
DT 01-MAR-2002 (TREMblrel. 20, Created)
DT 01-MAR-2002 (TREMblrel. 20, last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, last annotation update)
DE Similar to MEGF10 protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=MUSCLE;
RA Strausberg R.;
RL Submitted (DEC-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; BC020198; AAH20198.1; -.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR002049; Laminin_EGF.
DR Pfam: PF00008; EGF_9.
DR PRINTS: PR00011; EGF_LAMININ.
DR SMART; SM00181; EGF_11.
DR SMART; SM00180; EGF_Lam; 9.
DR PROSITE; PS00022; EGF_1; UNKNOWN_10.
DR PROSITE; PS01186; EGF_2; UNKNOWN_10.
SQ SEQUENCE 567 AA; 60797 MW; CFEFBGDEB7C627 CRC64;

Query Match 96.3%; Score 3468; DB 4; Length 567;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 565; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

Oy 1 MYISLNSCSFLCLCHITGASPLNEDPVNCSHMEYSTVYQDSYHPEDQIYITSC 60
Db 1 MYISLNSCSFLCLCHITGASPLNEDPVNCSHMEYSTVYQDSYHPEDQIYITSC 60
Oy 61 TPIILMFKCTRRHVSRYTRAYRHGEKTMRYRKSQCCPGFYESEEMCVPHCADRCVHGRCTA 120
Db 61 TPIILMFKCTRRHVSRYTRAYRHGEKTMRYRKSQCCPGFYESEEMCVPHCADRCVHGRCTA 120

```

```

Db 61 TDILMWFCTRRHVSRYTRAYRHGEKTMRYRKSQCCPGFYESEEMCVPHCADRCVHGRCTA 120
Oy 121 PNTCCPEPMGGTNCSSACDGDHMGPHCTSRCCCKNGALCNPTTGACHCAAGFRGMRCD 180
Db 121 PNTCCPEPMGGTNCSSACDGDHMGPHCTSRCCCKNGALCNPTTGACHCAAGFRGMRCD 180
Oy 181 RCEOGTYGNDCHQRCQCCNGATCDHYTGECRCRPPGYTGAFCEDLCPGKHGPOCEQRCP 240
Db 181 RCEOGTYGNDCHQRCQCCNGATCDHYTGECRCRPPGYTGAFCEDLCPGKHGPOCEQRCP 240
Oy 241 QNGGVCHHTTGBCSCPSGSMGTVCQPCPEGRGRKNCSEBCCCHNGTCDATGQCHCSP 300
Db 241 QNGGVCHHTTGBCSCPSGSMGTVCQPCPEGRGRKNCSEBCCCHNGTCDATGQCHCSP 300
Oy 301 GTTGECODECPVGTGYVCAETCCQVNGKCYHYSGALCEAGFAGRCERLCEPGLY 360
Db 301 GTTGECODECPVGTGYVCAETCCQVNGKCYHYSGALCEAGFAGRCERLCEPGLY 360
Oy 361 GIKCDKRCPCHEHTHSHCHPMSEGCACRPGMSGLYCNETCSPGFYGEACQOITSCQNGAD 420
Db 361 GIKCDKRCPCHEHTHSHCHPMSEGCACRPGMSGLYCNETCSPGFYGEACQOITSCQNGAD 420
Oy 421 CDSYTKCTCAGPFGKIDSTPCPLGTGYGINSRCCKNDVAVCSYVDSCTCKAGMHGV 480
Db 421 CDSYTKCTCAGPFGKIDSTPCPLGTGYGINSRCCKNDVAVCSYVDSCTCKAGMHGV 480
Oy 481 DCSIRPSTGTWFGCNLTQCCLNGACNTLDGTCTCAPMRGEKCELPQODTYGGLNCAE 540
Db 481 DCSIRPSTGTWFGCNLTQCCLNGACNTLDGTCTCAPMRGEKCELPQODTYGGLNCAE 540
Oy 541 RDCSHADCGHPTTGHCRLPGMSGY 566
Db 541 RDCSHADCGHPTTGHCRLPGMSGL 566

RESULT 3
O96KG6 PRELIMINARY: PRT: 969 AA.
ID O96KG6
AC O96KG6:
DT 01-DEC-2001 (TREMblrel. 19, Created)
DT 01-DEC-2001 (TREMblrel. 19, last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, last annotation update)
DE The complete sequences of 100 new cDNA clones from brain which code
DE for large proteins in vitro.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA MEDLINE-21245130; PubMed-11347906;
RA Nagase T., Nakayama M., Nakajima D., Kikuno R., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XX.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro."
RL DNA Res. 8:85-95(2001).
RL EMBL; AB058677; BAB47410.1; -.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001281; Rleske.
DR Pfam: PF00008; EGF_15.
DR PROSITE; PS00022; EGF_1; UNKNOWN_17.
DR PROSITE; PS01186; EGF_2; UNKNOWN_17.
DR PROSITE; PS00200; RIESKE_2; UNKNOWN_1.
SQ SEQUENCE 969 AA; 101600 MW; 56DDJFEEL39C8209 CRC64;

Query Match 58.2%; Score 2094; DB 4; Length 969;
Best Local Similarity 65.4%; Pred. No. 2,2e-194;
Matches 312; Conservative 58; Mismatches 107; Indels 0; Gaps 0;

```

```

Oy 109 CADKCVHGRICAPNTCCPEPMGGTNCSSACDGDHMGPHCTSRCCCKNGALCNPTTGACHCAAGFRGMRCD 168
Db 28 CTBEVYHGRCVSPDITCHEPFGMGPPDCSSGCDSDHMGPHCTSRCCCKNGALCNPTTGACHCAAGFRGMRCD 87

```

```
OY 169 CAAGFRGRCEDRCDEGNYGNDCHQROCONGATCDHVTGECRCRPGYTAFCEDLCPG 228
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 88 CAAGFRGRCEDLCPGTHGKCOLPCQCRHAGASCDDPRAGCLAPGTGYCELCPPG 147
OY 229 KHGRCORRCPCONGVCHHTGECSCPSGMMGTVCQPCPEGRFGKNSQEOCHNGT 288
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 148 SHGACELRCPCONGGCHHTGECACRPGMTGAVCAQPCRPRTFGQCSDDCPCHGQ 207
OY 289 CDATGCHCSPGTYTGERCDECPYGYTVCACETCCQVNGCKYHSGACLCEGRAGE 348
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 208 CDHATGCHCAGTYMGRCOECPFGSCQRCDCDHNGQCSPTTGACCECEGXYGP 267
OY 349 RCEARLCPGLYGIKCRKCPCHLEHTHSCHPMSGECACRKGSLYCNETSCEGFEYA 408
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 268 RQGRLECPGLHGPGLLPCPCDADNTISCHPVTGACTCQGMGSHHCNESCPRYIGDG 327
OY 409 CQOICSCONGACDSDVTGKCTCAPGFGIDSTPCPLGTGYNCSRCKNDAYCSPVD 468
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 328 COLPCTCONGADSHITGGCTCAGPFGMEYCAVSCAGTGPCNCSITSCNNGTCSFVD 387
OY 469 GSCCTKAGMHVDCSIRGSPSTNGFCNLTQCLNGACNTLDITCIPAGMREKCELP 528
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 388 GSCCTKAGMHVDCSIRGSPSTNGFCNLTQCLNGACNTLDITCIPAGMREKCELP 447
OY 529 CODGTGYNCAERCDSCSHADGCHPTTGCHRCPLPGMSGVHCDSCVCAEGRMGPCSLPC 585
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 448 CPDPTGFLNGSEHRCDCSHADGCDPVTGHCCLAGMTGIRKCDSTCPPRGMGNCVSC 504
```

RESULT 4

08VHF4 PRELIMINARY; PRT; 747 AA.

```
ID 08VHF4
AC 08VHF4
DT 01-MAR-2002 (Tremblrel, 20, Created)
DT 01-MAR-2002 (Tremblrel, 20, Last sequence update)
DT 01-JUN-2002 (Tremblrel, 21, Last annotation update)
DE Jed1-736 protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL; TISSUE=TESTIS;
RA Krivosov A.V., Zinoviyeva M.V., Hendrikx J., Visser J.W.M.,
Belyavsky A.V.;
RT "Jed1 is a novel DSL and EGF-like repeat motif-containing protein
expressed on non-differentiated hematopoietic cells."
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF461685; AAL66380.1; -
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR002049; Laminin_EGF.
DR Pfam: PF00008; EGF_9.
DR PRINTS: PR00011; EGF_LAMININ.
DR SMART: SM00181; EGF_15.
DR SMART: SM00180; EGF_Lam; 14.
DR PROSITE: PS00022; EGF_1; UNKNOWN_13.
DR PROSITE: PS01186; EGF_2; UNKNOWN_12.
DR PROSITE: PS01186; EGF_2; UNKNOWN_12.
SQ SEQUENCE 747 AA; 78972 MW; F825F8F84D4736A CRC64;
```

Query Match 50.8%; Score 1828; DB 11; Length 747;
Best Local Similarity 52.0%; Pred. No. 1e-168;
Matches 299; Conservative 59; Mismatches 211; Indels 6; Gaps 3;

```
OY 14 LLLCHWIGTASPLNLEDPNVCSHWESYVYQESYRPHFDIYYTCTDILNM--FKCT 70
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 7 LLLALGLRLGTLSNDPNCTFESFTTTTKESHLLRPFSLPAESCH--RPWEDPHCA 64
OY 71 RHRYSYTAAYRHGEKTYMRKRSQCCPGFYESGEMCVPHCADKCYHGRCIAPNTQCEBPG 130
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 65 QPTVYVRYVRYQVVKMDSRPLQCCRGYTESRGACVPLCAQECVHGRCVAPNOCOCAPGW 124
```

```
OY 131 GGTNCSSACDHDHNGPHCTSRQCKNGALCNPIITGACHCAAGFRGMRCEDECRQTYGND 190
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 125 RGDCSSSEACAGMMGPOCDKFCRCHGNSSCCKPSKACCPGSLDPPNCLDPCRGHGPRA 184
OY 191 CHOROCONGATCDHVTGECRCRPGYTAFCEDLCPGKHPOGECORCPCONGVCHHT 250
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 185 CQFQCCGY-GASCDPQDACCPCPRBAGPSCNVPCSQGTDFCFPRYPCQNGVPGQSQ 243
OY 251 GECSCPGSMKMTVCQPCPEGRFGKNSQEOCHNGTCDATATQCHSPGTYGERCODE 310
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 244 GSCSCPFGMMVITSLPCPEGFHGPNTQECRCHNGGLCDFTQCHCAPGYIDRCOE 303
OY 311 CPVGTGYVLCATQCCQVNGKCYHSGACLCEAGFAGRCARLCPGLYGIKCDKRCPC 370
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 304 CPVGRFQDCAETDCABAGANCIPFANGACLCHEHGTGDRCTERLCIPDRYTLSCQEPRTC 363
OY 371 HLENTHSCHPNSGECACRKGSLYCNETSCEGFEYGEACQOICSCONGACDSDVTGKCTC 430
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 364 DPEHSLSCHPMHGECSQPGWAGLHCNESCQPDTHGPGCQDHCLCLHGLCLADSLGRC 423
OY 431 APGFKGIDCTSPCLGTGYNCSRCKNDAYCSPVDGSCCTKAGMHVDCSIRP8GT 490
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 424 APGTYGPRCANLCPDPTGYNCSRCKNDAYCSPVDGSCCTKAGMHVDCSIRP8GT 483
OY 491 WGFPCNLTCQCLNGACNTLDGTCAPGMRGKCELPQDGTGYNCAERCDSCSHADGC 550
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 484 WGFPCNLTCQCLNGACNTLDGTCAPGMRGKCELPQDGTGYNCAERCDSCSHADGC 543
OY 551 HPTTGCHRCLEPGMSGVHCDSCVCAEGRMGPCSLPC 585
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 544 DPHGQCRQAGQMMGTGRCHLCPREGFWGANCSTNC 578
```

RESULT 5

08VHF7 PRELIMINARY; PRT; 1034 AA.

```
ID 08VHF7
AC 08VHF7
DT 01-MAR-2002 (Tremblrel, 20, Created)
DT 01-MAR-2002 (Tremblrel, 20, Last sequence update)
DT 01-JUN-2002 (Tremblrel, 21, Last annotation update)
DE Jed1 protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL; TISSUE=TESTIS;
RA Krivosov A.V., Zinoviyeva M.V., Hendrikx J., Visser J.W.M.,
Belyavsky A.V.;
RT "Jed1 is a novel DSL and EGF-like repeat motif-containing protein
expressed on non-differentiated hematopoietic cells."
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF444274; AAL38571.1; -
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR002049; Laminin_EGF.
DR Pfam: PF00008; EGF_9.
DR PRINTS: PR00011; EGF_LAMININ.
DR SMART: SM00181; EGF_15.
DR SMART: SM00180; EGF_Lam; 14.
DR PROSITE: PS00022; EGF_1; UNKNOWN_13.
DR PROSITE: PS01186; EGF_2; UNKNOWN_12.
DR PROSITE: PS01186; EGF_2; UNKNOWN_12.
SQ SEQUENCE 1034 AA; 110540 MW; 5514E3166AE01111 CRC64;
```

Query Match 50.8%; Score 1828; DB 11; Length 1034;
Best Local Similarity 52.0%; Pred. No. 1.4e-168;
Matches 299; Conservative 59; Mismatches 211; Indels 6; Gaps 3;

```
OY 14 LLLCHWIGTASPLNLEDPNVCSHWESYVYQESYRPHFDIYYTCTDILNM--FKCT 70
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 7 LLLALGLRLGTLSNDPNCTFESFTTTTKESHLLRPFSLPAESCH--RPWEDPHCA 64
OY 71 RHRYSYTAAYRHGEKTYMRKRSQCCPGFYESGEMCVPHCADKCYHGRCIAPNTQCEBPG 130
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
```

Db	65	QPTVYRIVYQVYVKMDSRPRLQCCRGTYEBSRGACVPLCAQECVHGRCVAPRQCCSACGW	12
QY	131	GGTNCSSACDGDHMGPHCTSRQCCNGALCNPTGACCAAGFRGWRCEDRCEQTYGND	19
Db	125	RGGDCSSCAGGMMGPGQDKFCFHCGNNSSDCPKSGACFCPSGLDPRNCLDPRAGHYGRA	18
QY	191	CHORQCCNGATCDHMYBECNCPREYTYAFEDLCPRKHNQRCQRCRQCCNGYCHNVT	25
Db	185	COFDQCY-GASCDPQDQACFCSPRGARPCSNVPCSGQTDGFPCRTYPCQNGVPGSQ	24
QY	251	GECSGPSGMMGTVCQGPPEGRFGRKNSQEOCCNHGTCDAATGQCHSCSPGYTGERCODE	31
Db	244	GSCSCPPEMMVIGSLPRPGRFHNKNTQOEBRCRHNGGLCDFTGQCHAPETIDRCQBE	30
QY	311	CPVGTGYVLCAETCCQVNGKRCYHVSACILCEAGFAGERCEARLCPBGLYGIKDKRCP	37
Db	304	CPVGFQGDCAETCCAGARCFRANGACILCBHGTDRCTERLCPDGRYFGHSCDEPCTC	36
QY	371	HLENTHSHRPSBEACRPRHSGILCNFTCSBGFGEAGCQDLCSCQNAQDSDVYKCTC	43
Db	364	DREHLSCHPHNGECSRPGMAGLHCNCSQPDTHNGPCCQHEICLCLHGLCLALASGLRC	42
QY	431	ARFGKIGCSTPRCLPTGYGKINCSHRGCKNDVAVCSPLYVGSCTCKRAGMAYDQCSIRCPSGT	49
Db	424	ARGYGRPCANLCPRPDTYTGKINSSRSCSENNIACSRIDGTCICKRGMQRKMSVYCPPLGT	48
QY	491	WFGCNLTCCQCLNGACNTLDGTCTCABRWGKEKELPCQDGTGVLNCAERDCSHADQC	55
Db	484	WGFNCAACQCAHNDVCSPTQGTACTCTGPMGHANQCLPCPKQGFEGCASVQDCDHSDQC	54
QY	551	HPYTHCRKRLPQMSGVHDSVCABEWRMPCNSLPC	55
Db	544	DVYHQCRCQAGMMGTCHLPCPRBEGFWANCSTNC	578

```

RESULT 6
08VITS      ID          Q8VIK5      PRELIMINARY;    PRT; 1034 AA.
AC           Q8VIK5;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE MEGF12.
GN MEGF12.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10990;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=LIVER;
RA Ivanova N.B., Lemischka I.R.;
RT "The global gene expression profiling of the hematopoietic stem
   cell".
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF440279; AAL33583.1; -
DR InterPro; IPRO00561; EGF-like.
DR InterPro; IPRO02049; Laminin_EGF.
DR Pfam; PF00008; EGF; 9.
DR PRINTS; PRO0011; EGFLAMININ.
DR SMART; SM00181; EGF; 15.
DR SMART; SM00180; EGF_Lam; 14.
DR PROSITE; PS00022; EGF_1; UNKNOWN_13.
DR PROSITE; PS01186; EGF_2; UNKNOWN_12.
DR SEQUENCE 1034 AA; 110580 MW; 714E501684BEA4C CRC64;

Query Match      50.7%; Score 1824; DB 11; Length 1034;
Best Local Similarity 51.8%; Pred. No.3.4e+166;
Matches 298; Conservative 59; Mismatches 212; Indels 6; Gaps 3.

```

Db	7	LLALGLRLTGLTSLNSNDPNNCTFHESTPTTTT KESHLRPSFSLPAESCH -- RPWEDPHITCA	64
Qy	71	RHRVSYRTAYNHGEKTYMRRKSQCCPGFTYBSGEMCVPHADCKVHGKFLAPNTQCEPBGW	130
Db	65	QPTVYRTVYRQVYVKMDSRPRLQCCRCRGYTESRGACVPLCAQCEVHGKCVAPNCCOCAPGW	124
Qy	131	GGTNCSSACDSDHMGPHCTSHSQCKNKALCNPTGACNCAAGFRGMGCEDECRGGYGN	190
Db	125	RGGDSSSCACGMMGPPCQDKFCRCHGNSSDCPSGTCFCPSGLQPPNCLPCPAGHGYPA	184
Qy	191	CHORCKONGATCTCHVTGECRCRPGYTGACEDLCPPGKHGRPOECRCPQNGVCHNYT	250
Db	185	CGPFOGQC - GASCDPDQDAGACCPGGRAGRPSCNVPSCSGTGTFPCPRRYPCQNGVYQGSQ	243
Qy	251	GECSCPSGMGTGVCGRPCPEGRFGKNSQSEQCNGSTGCTAATGQCHSCPFTYGERCODE	310
Db	244	GSCSCPSPMMVYLSLPCPEGFHGNPTQTECRCHNGSLCDRFTGQCHACAPGYSGRDQEE	303
Qy	311	CPVGTGYVLCAETQCCVNGGKCYHVSAGCLCEAGFAEREARLCRPGGLYGTICDKRCP	370
Db	304	CPVAFRGQDCAETEDCDAPGACCFPANCACLCHEFTEDRCTERLCRPGRGLSCQEECTC	353
Qy	371	HLENTYCHPMSGECACKPGSGSLYICNETDSPPGYEGACQICSCQNGACDSDVTGKTC	430
Db	364	DPEHSLSCHPHHGECSQCPGNAIGHCHESCPDQTHHGECQDHLCLHGGICLADSLGLCR	423
Qy	431	APGFKGIDCSYPCPLGTGYINGICSSRCGCKNDAYCSPYDSCSTCKAGNHGYDCSIRCPSGT	490
Db	424	APGTGPHCANLCPRPDTYIGICSSRCSEENAIACSPIDGTCICKEGQRQNSVYPLGT	483
Qy	491	WFGCNLTQCLNGACNTLDGTCTCAAPGRGKEKCELPQDDGYTGLNCAERCDCHADGC	550
Db	484	WGFNMNASCQCAHGDVCSPTQAGACTCTPGMHGAHCOLPCPRGQFGEBCASVCDCHSDGC	543
Qy	551	HTTGHCRCLRGMGSHVDSCYCAGRGGRPNLSLPC	585
Db	544	DVPHGQCRQAGMMGTGRCHLPCPGTGFAGNCSNTC	578

```

RESULT 7
088281
ID      088281      PRELIMINARY;      PRT;      1574      AA.
AC      088281;
DT      01-NOV-1998      (TrEMBLrel. 08, Created)
DT      01-NOV-1998      (TrEMBLrel. 08, Last sequence update)
DT      01-DEC-2001      (TrEMBLrel. 19, Last annotation update)
DE      ME6F6.
GN      ME6F6.
OS      Rattus norvegicus (Rat).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX      NCBI_TaxID=10116;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=SPRAGUE-DAWLEY; TISSUE=BRAIN;
RX      MEDLINE=98360089; PubMed=9693030;
RA      Nakayama M., Nakajima D., Nagase T., Nomura N., Seki N., Ohara O.;
RT      "Identification of high-molecular-weight proteins with multiple EGF-
RT      like motifs by motif-trap screening.";
RL      Genomics 51:27-34(1998).
DR      EMBL; AB011532; BAA32462.1; -.
DR      HSSP; P00736; IAP0.
DR      InterPro: IPR000152; Asx_hydroxyl.
DR      InterPro: IPR000561; EGF-like.
DR      InterPro: IPR001881; EGF_Ca.
DR      Pfam; PF00008; EGF_24.
DR      SMART; SM00179; EGF_Ca; 4.
DR      SMART; SM00001; EGF_like; 19.
DR      PROSITE; PS00010; ASX_HYDROXYL; 5.
DR      PROSITE; PS00022; EGF_1; UNKNOWN_23.
DR      PROSITE; PS01166; EGF_2; 23.
DR      PROSITE; PS01187; EGF_Ca; 5.
DR      Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat
FW

```


SO SEQUENCE 1574 AA; 165445 MW; 2848533DBF77FE67 CRC64;

Query Match 38.1%; Score 1372.5; DB 11; Length 1574;

Best Local Similarity 41.3%; Pred. No. 2.8e-124; Matches 229; Conservative 56; Mismatches 198; Indels 71; Gaps 9;

QY 94 CCEPGESEGMCVHACDKVHG-----RCIAPN-----"TQCEPGMGGINCS 136
 DB 808 CLPFGVS-----RCODTCSAGMYTGCCQIRACACANDGCHDPTTGKSCAPRGMTGLSCQ 861
 QY 137 SAGDGHMGPHTSCRCCKNG-ALCNDPTGACACAGFRGWRCEDEQTYGNDCHQRC 195
 DB 862 RACDSGHMGPDCIHPNCMSAGHGNCDASVGLCEAGYEGRCBQSGRGYGGSCBQKC 921
 QY 196 QCCGATCDHYTGRCRPPGTGAFCD-----LC 225
 DB 922 RCEHGAACDHVSAGCTCPAGWRGSCFHACPAGEFGLDCDSACNCSAGAPCDAYTGSIC 981
 QY 226 PGRKHGQCEBQRCP-----CONGVCHHTGECSCPSGMWGTVCQGPCEGR 272
 DB 982 PAGRWGRCAQSCPTLFGNLCSQICTFNASCDSVTGQCHCAPGMWPTCLDACPPL 1041
 QY 273 FGNKSCQECCHNGTCDAAATGQCHSPGYTGERCODECPVGTGVLCAETQCQCVNGKC 332
 DB 1042 YGKNCQHSCLCRNGRCDDPLIGQCTCEGWTGLACENECILPGHYAAGCQLNCSCLHGIC 1101
 QY 333 YHVSAGALCEAGFRGCEARLCEGLXGKDKRCRCHLENTHSCHPMSECCAKRGWS 392
 DB 1102 DRLTGHCLCPAGMTGDCQSS-CVSGTFGVHCEHCAC--RKGASCHHTVGCPCPGWR 1158
 QY 393 GLYCNETSPGFYEACQQLCSQNGADCSVTGKCTCAFCFKGIDSTPCPLGTGIC 452
 DB 1159 GPHQEQACPRGWRFEACAQRCLEPTNSCHHTVGEBCRCPGFTLSBQACOPRTFKDC 1218
 QY 453 SSRGCGKNDV-VCSPVDGCTCAAGHGVDCSIRCPSGTWGFGCNLTQCLNGACNTLD 511
 DB 1219 EHLQCCGEGTACDPASGVTCTCAAGYGTGCLQRCPSGRGRCHEHCKLNGSTCDPAT 1278
 QY 512 GTCTCAGWNGEKEKELPCODGTGVLNCAERDCSHADGCHPTTGCHCLPMSGVHDSV 571
 DB 1279 GACTCPRGFLGADCSLCPGGRFPSCAHYACRQAGACDPSGACICSPKGTVCBEHG 1338
 QY 572 CAEGRMGPNCLPC 585
 DB 1339 CPQDRFGKCELC 1352

RESULT 8
 ID 09TV02 PRELIMINARY; PRT; 1664 AA.
 AC 09TV02;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE Y64G10A.7 protein.
 GN Y64G10A.7.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Peloderiinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN SEQUENCE FROM N.A.
 RA Mortimore B.J.;
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
 RP [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99069613; PubMed=9851916;
 RA none;
 RT "Genome sequence of the nematode C.elegans: A platform for
 RL Investigating biology".
 RL Science 282:2012-2018(1998).
 RP [3]
 RP SEQUENCE FROM N.A.

RA Ainscough R.;
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL; AL117206; CAB60454.1; JOINED.
 DR EMBL; AL110498; CAB60454.1; JOINED.
 DR EMBL; AL110498; CAB57911.1; JOINED.
 DR EMBL; AL117206; CAB57911.1; JOINED.
 DR HSSP; P00736; JAP0.
 DR InterPro; IPR001052; Asx_hydroxyl.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR002049; Laminin_EGF.
 DR Pfam; PF00008; EGF; 25.
 DR PRINTS; PR00011; EGF_LAMININ.
 DR SMART; SM00179; EGF_CA; 4.
 DR SMART; SM00001; EGF_Like; 18.
 DR PROSITE; PS00010; ASX_HYDROXYL; 4.
 DR PROSITE; PS00022; EGF_1; UNKNOWN_22.
 DR PROSITE; PS01186; EGF_2; 24.
 DR PROSITE; PS01187; EGF_CA; 3.
 DR Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.
 SO SEQUENCE 1664 AA; 179279 MW; A69F093B4C705832 CRC64;

Query Match 37.2%; Score 1340; DB 5; Length 1664;
 Best Local Similarity 36.4%; Pred. No. 4.2e-121;
 Matches 237; Conservative 70; Mismatches 218; Indels 126; Gaps 13;

QY 16 LCHNI-----GTASPL-----NLEDPNCSHNESTVYQESYHPHPDQIYTSC 60
 DB 824 VCHHTVCTCLPCKTLPCLDQSCAPNTYGPN-CAH-----TC 860
 QY 61 TDILNFKCTRHRVSYTAYRHGEKTYRRKSQCCPGEYSEGMCVPHCAD-----TC 111
 DB 861 S-CYNAKCESDS-----CHCTPPTFY--GATCSEVCTPGRGIDCMQ 901
 QY 112 --KCVHR-CLAPN-TQCEPGWGTNCSSACDGDHMGPHCTSCQCKNALCNPTGAC 167
 DB 902 LCKQNGAICDTSNGSECECAPAGSGKRCDAAPGTGPKDCSKKCDACADMHCDPDSGEC 961
 QY 168 HCAAGFGMGCEDCEDEGTGNDCHQCCONGAT----- 202
 DB 962 ICPGKRGKHCDCTGDSGLFAGAGCKGICSCQNGATCDSTVSGSCRPGRGKKCDRCPD 1021
 QY 203 -----CDHVTGECRCPPGYTGAFCECLCPGKHGPOC 234
 DB 1022 GRREGCNALCDCTTNDTSMYNPFAVARCHDHYTECCPAGWTGPDQTSCLPRHGECC 1081
 QY 235 EQRCPQNGGVCHHTGECSCPSGMWGTVCQPCPEGRFGRKNSQECQCHNGTCDATG 294
 DB 1082 RHSCQCSNGASCDREVTFCDPCPSGFMKNCESCEPGLMGSNCMKHCLCMHNGECNKENG 1141
 QY 295 QCHSPGYTGERCODECPVGTGVLCAETQCQCVNAGGCIYHSGACLCFAFGRCARL 354
 DB 1142 DCEIDTDMTPSECEFLCPFOQFGRNCAORCNCKNGASCADRKTGCECLPMSGHECH-KS 1200
 QY 355 CPEGLYGIKDKRCPCHEMTHSCHPMSECCAPRGSGLYGNCNTGSCPGFYGEACQICS 414
 DB 1201 CVSHYAKCEBTECC--ENGALCDPLISGHCSCQPGKRGKCNRPCLAKGFRGHCSSCR 1258
 QY 415 CQNGADCSVTGKCTCAFGKIDCSTPCPLGTGVLGNCSSRCGCKNDVAVSPVDGSCYCK 474
 DB 1259 CANSKSCDHISGRQCPKGYAGHSCFTELCPDGFGECSQCKDCGEGNSMDAISGKCFCK 1318
 QY 475 AGHGVDCSIRCPSGTGFEGCNLTQCLNGACNTLDGTCTCAPGMWGEKELPCODGT 534
 DB 1319 PGHSGSDCKSGGVQGRGPCCNOLCSCENGVCDSSTGSCVCPGPGYGTGCELAQSDRF 1378
 QY 535 GLNCAERDCSHADGCHPTTGCHCLPMSGVHDSVCAEGRMGPNCLPC 585
 DB 1379 GPICERKICENGSTCDRLTGQCRCLPFTGMTCNQCVPBGRGAGCKENC 1429

RESULT 9
 ID 063404

```

ID 063404 PRELIMINARY: PRT: 220 AA.
AC 063404:
RX MEDLINE-99069613; PubMed-9851916;
RT "Genome sequence of the nematode C. elegans: A platform for
  investigating biology."
RL Science 282:2012-2018(1998).
DR EMBL; AL032657; CAD27615.1;
SQ SEQUENCE 1045 AA; 111723 MW; 0A6D57A3A80BCDCA CRC64;

Query Match
Best Local Similarity 35.7%; Score 1284.5; DB 5; Length 1045;
Matches 246; Conservative 77; Mismatches 221; Indels 175; Gaps 20;

OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OC NCBI_TaxId:10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-HOLZMAN; TISSUE-BRAIN;
RX MEDLINE-9623515; PubMed-8642059;
RA Asakura K., Pogulis R.J., Pease L.R., Rodriguez M.;
RT "A monoclonal antibody which promotes central nervous system
  remyelination is highly polyreactive to multiple known and novel
  antigens."
RL J. Neuroimmunol. 65:11-19(1996).
DR EMBL; L41686; AAB05844.1;
RX HSSP; P01132; 1EGF.
DR InterPro: IPR000561; EGF_Like.
DR InterPro: IPR002049; Laminin_EGF.
DR Pfam: PF00008; EGF_3.
DR PRINTS; PR00011; EGF_LAMININ.
DR SMART; SM00180; EGF_Lam; 2.
DR SMART; PS00022; EGF_Like; 1.
DR PROSITE; PS01186; EGF_2; 5.
DR EGF-like domain; Glycoprotein.
KW NON_TER
FT NON_TER 1
FT NON_TER 220
SQ SEQUENCE 220 AA; 23231 MW; 3119D391EAF64372 CRC64;

Query Match
Best Local Similarity 36.4%; Score 1309; DB 11; Length 220;
Matches 209; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

OY 200 GATCCHVTGECRCPPGYTGAFCECLCPGKHGPCEQORPCPCONGCVCHVTGECSPSGM 259
DB 2 GATCCHVTGECRCPPGYTGAFCECLCPGKHGPCEQORPCPCONGCVCHVTGECSPSGM 61
OY 260 MCTVCGOPCEPRFGKNCSEOCCHNGTCDATGQCCHCSPGYTGRCODECPVGTGYVL 319
DB 62 MCTVCGOPCEPRFGKNCSEOCCHNGTCDATGQCCHCSPGYTGRCODECPVGTGYVL 121
OY 320 CAETCOCVNGGKCYHVSACLCGAPGRCERARLCPSGLYGIKDKRCPCHLENTHSCH 379
DB 122 CAETRCVNGGKCYHVSACLCGAPGRCERARLCPSGLYGIKDKRCPCHLENTHSCH 181
OY 380 PMSGECACRPGMSGLYCNETCSPGYGACQOICSCONG 418
DB 182 PMSGECACRPGMSGLYCNETCSPGYGACQOICSCONG 220

RESULT 10
OBT3A6 PRELIMINARY: PRT: 1045 AA.
AC 08T3A6:
RX MEDLINE-99069613; PubMed-9851916;
RT "Genome sequence of the nematode C. elegans: A platform for
  investigating biology."
RL Science 282:2012-2018(1998).
DR EMBL; AL032657; CAD27615.1;
SQ SEQUENCE 1045 AA; 111723 MW; 0A6D57A3A80BCDCA CRC64;

OY 200 GATCCHVTGECRCPPGYTGAFCECLCPGKHGPCEQORPCPCONGCVCHVTGECSPSGM 259
DB 2 GATCCHVTGECRCPPGYTGAFCECLCPGKHGPCEQORPCPCONGCVCHVTGECSPSGM 61
OY 260 MCTVCGOPCEPRFGKNCSEOCCHNGTCDATGQCCHCSPGYTGRCODECPVGTGYVL 319
DB 62 MCTVCGOPCEPRFGKNCSEOCCHNGTCDATGQCCHCSPGYTGRCODECPVGTGYVL 121
OY 320 CAETCOCVNGGKCYHVSACLCGAPGRCERARLCPSGLYGIKDKRCPCHLENTHSCH 379
DB 122 CAETRCVNGGKCYHVSACLCGAPGRCERARLCPSGLYGIKDKRCPCHLENTHSCH 181
OY 380 PMSGECACRPGMSGLYCNETCSPGYGACQOICSCONG 418
DB 182 PMSGECACRPGMSGLYCNETCSPGYGACQOICSCONG 220

RESULT 11
OBT3A7 PRELIMINARY: PRT: 1070 AA.
AC 08T3A7:
RX MEDLINE-99069613; PubMed-9851916;
RT "Genome sequence of the nematode C. elegans: A platform for
  investigating biology."
RL Science 282:2012-2018(1998).
DR EMBL; AL032657; CAD27615.1;
SQ SEQUENCE 1045 AA; 111723 MW; 0A6D57A3A80BCDCA CRC64;

OY 200 GATCCHVTGECRCPPGYTGAFCECLCPGKHGPCEQORPCPCONGCVCHVTGECSPSGM 259
DB 2 GATCCHVTGECRCPPGYTGAFCECLCPGKHGPCEQORPCPCONGCVCHVTGECSPSGM 61
OY 260 MCTVCGOPCEPRFGKNCSEOCCHNGTCDATGQCCHCSPGYTGRCODECPVGTGYVL 319
DB 62 MCTVCGOPCEPRFGKNCSEOCCHNGTCDATGQCCHCSPGYTGRCODECPVGTGYVL 121
OY 320 CAETCOCVNGGKCYHVSACLCGAPGRCERARLCPSGLYGIKDKRCPCHLENTHSCH 379
DB 122 CAETRCVNGGKCYHVSACLCGAPGRCERARLCPSGLYGIKDKRCPCHLENTHSCH 181
OY 380 PMSGECACRPGMSGLYCNETCSPGYGACQOICSCONG 418
DB 182 PMSGECACRPGMSGLYCNETCSPGYGACQOICSCONG 220

RESULT 12
OBT3A8 PRELIMINARY: PRT: 1070 AA.
AC 08T3A8:
RX MEDLINE-99069613; PubMed-9851916;
RT "Genome sequence of the nematode C. elegans: A platform for
  investigating biology."
RL Science 282:2012-2018(1998).
DR EMBL; AL032657; CAD27615.1;
SQ SEQUENCE 1045 AA; 111723 MW; 0A6D57A3A80BCDCA CRC64;

OY 200 GATCCHVTGECRCPPGYTGAFCECLCPGKHGPCEQORPCPCONGCVCHVTGECSPSGM 259
DB 2 GATCCHVTGECRCPPGYTGAFCECLCPGKHGPCEQORPCPCONGCVCHVTGECSPSGM 61
OY 260 MCTVCGOPCEPRFGKNCSEOCCHNGTCDATGQCCHCSPGYTGRCODECPVGTGYVL 319
DB 62 MCTVCGOPCEPRFGKNCSEOCCHNGTCDATGQCCHCSPGYTGRCODECPVGTGYVL 121
OY 320 CAETCOCVNGGKCYHVSACLCGAPGRCERARLCPSGLYGIKDKRCPCHLENTHSCH 379
DB 122 CAETRCVNGGKCYHVSACLCGAPGRCERARLCPSGLYGIKDKRCPCHLENTHSCH 181
OY 380 PMSGECACRPGMSGLYCNETCSPGYGACQOICSCONG 418
DB 182 PMSGECACRPGMSGLYCNETCSPGYGACQOICSCONG 220

```

```

RP SEQUENCE FROM N.A.
RX MEDLINE-99069613; PubMed-9851916;
RA none;
RT "Genome sequence of the nematode C. elegans: A platform for
  investigating biology."
RL Science 282:2012-2018(1998).
DR EMBL; AL032657; CAD27615.1;
SQ SEQUENCE 1045 AA; 111723 MW; 0A6D57A3A80BCDCA CRC64;

Query Match
Best Local Similarity 34.2%; Score 1284.5; DB 5; Length 1045;
Matches 246; Conservative 77; Mismatches 221; Indels 175; Gaps 20;

OY 21 GTASLNLIEDPNVSCSHWESYVTVQESYRPHPDQIYTT-----SCDILNMFKSTRHR 73
DB 35 GTTEP---GQHVCV-----VKTIYDDY--ELKVIHYVYNDTQCLNPLPLGFCF---- 80
OY 74 VSYRTAIRHGEKTYMTRK-----SQCPCPFYSGE-MCVPHCADKCVHGRCTAPNTC 124
DB 81 ----TVERKGOXKASYORQLVKKKRYKOCDDQYOTKDHFLCDNPNPKKKGKCIPEPKC 136
OY 125 OCEPMGNTNCSACDGDHMGPHCTSRQCKNGALCNPTGCHCAAGRGRC----- 179
DB 137 ECDPGLGKRYCASCSVGTWGLGCSKSCDCENGANCDEPLGTCITSGFQGERCKPCPD 196
OY 180 -----DRCEGTGYNDCRORCOCONGAT 202
DB 197 NKGPNVCVKSQCQNGKCKNGKGCVCSDMGWGFECILNKCEGKRGABECKFCNCNGAT 256
OY 203 CDHVTGECRCPPGYTGAFCECLCPGKHGPCEQORPCPCONGCVCHVTGECSPSGMGT 262
DB 257 CDNTNGKCIKCKSGYHAGALCENECISVGFSGGCTOKCDCILNNOCCSSSECCGC-IGWTGK 315
OY 263 VCGOPCEPRFGKNCSEOCQ-----HNGTCDATGQCCHCSPGYTGRCODECPVGT 315
DB 316 HCDIGSKRERFLQCKQNTCTCPGLEPSDSNACDANKTGCQCESGYKGGKCDERKCDAEQ 375
OY 316 YGVLCATETCOCY--NGKCYHVSACLCGAPGRCERARLCPSGLYGIKDKRCPCHLE 373
DB 376 YGADSKTCTCYRENTLMCAPRTGFCRCRKPGRYGNCEL-ACSKSYGNCKKQAMCDWN 434
OY 374 NTHSCHPMSGECACRPGMSGLYCNETCSPGYGACQOICSC-QNGADCDSTYGTCTCAP 432
DB 435 HASECNPEPESVCKPRGRKNCSEPCPLDFGPNCAHCCQCNQGVGDDGADGACQCCDR 494
OY 433 GFKGIDCSNPCLGTGYGICSSRCGCKNDAYVSPVDSGSCACAGHWYDCSTRCPSTWG 492
DB 493 GWTGHRCEHCHPADTFGANCERKCKCPKIGCDPITGECTCPAGLQGANCDIGCEBGSIG 554
OY 493 FGCNLTCCOLNAGACNTLDTCTCAPG-----RGEKCEL--PCQD----- 531
DB 555 PCCKLHKCVN-GKCDKRTGECTCQPGFPGSDCTTCSKGRGECCELSPCSDASCSKO 613
OY 532 -----GTGLNCAERCD----- 543
DB 614 TGKCLCLPLTKGVSQDCKDPTMFGFLCOEYVTPSPCASTDPRKNCVCLSCPSSGSHCE 673
OY 544 -----CSHAD--GCHPTTGHCRLCPGMSVHCHDSVACGRMPCSLPC 585
DB 674 HNCPRGSGTGDCCQVYCSADHGHCPTTGECICEFGTHGKTCSEKCPDGKIGYGCALDC 732

RESULT 11
OBT3A7 PRELIMINARY: PRT: 1070 AA.
AC 08T3A7:
RX MEDLINE-99069613; PubMed-9851916;
RT "Genome sequence of the nematode C. elegans: A platform for
  investigating biology."
RL Science 282:2012-2018(1998).
DR EMBL; AL032657; CAD27615.1;
SQ SEQUENCE 1045 AA; 111723 MW; 0A6D57A3A80BCDCA CRC64;

OY 200 GATCCHVTGECRCPPGYTGAFCECLCPGKHGPCEQORPCPCONGCVCHVTGECSPSGM 259
DB 2 GATCCHVTGECRCPPGYTGAFCECLCPGKHGPCEQORPCPCONGCVCHVTGECSPSGM 61
OY 260 MCTVCGOPCEPRFGKNCSEOCCHNGTCDATGQCCHCSPGYTGRCODECPVGTGYVL 319
DB 62 MCTVCGOPCEPRFGKNCSEOCCHNGTCDATGQCCHCSPGYTGRCODECPVGTGYVL 121
OY 320 CAETCOCVNGGKCYHVSACLCGAPGRCERARLCPSGLYGIKDKRCPCHLENTHSCH 379
DB 122 CAETRCVNGGKCYHVSACLCGAPGRCERARLCPSGLYGIKDKRCPCHLENTHSCH 181
OY 380 PMSGECACRPGMSGLYCNETCSPGYGACQOICSCONG 418
DB 182 PMSGECACRPGMSGLYCNETCSPGYGACQOICSCONG 220

RESULT 12
OBT3A8 PRELIMINARY: PRT: 1070 AA.
AC 08T3A8:
RX MEDLINE-99069613; PubMed-9851916;
RT "Genome sequence of the nematode C. elegans: A platform for
  investigating biology."
RL Science 282:2012-2018(1998).
DR EMBL; AL032657; CAD27615.1;
SQ SEQUENCE 1045 AA; 111723 MW; 0A6D57A3A80BCDCA CRC64;

OY 200 GATCCHVTGECRCPPGYTGAFCECLCPGKHGPCEQORPCPCONGCVCHVTGECSPSGM 259
DB 2 GATCCHVTGECRCPPGYTGAFCECLCPGKHGPCEQORPCPCONGCVCHVTGECSPSGM 61
OY 260 MCTVCGOPCEPRFGKNCSEOCCHNGTCDATGQCCHCSPGYTGRCODECPVGTGYVL 319
DB 62 MCTVCGOPCEPRFGKNCSEOCCHNGTCDATGQCCHCSPGYTGRCODECPVGTGYVL 121
OY 320 CAETCOCVNGGKCYHVSACLCGAPGRCERARLCPSGLYGIKDKRCPCHLENTHSCH 379
DB 122 CAETRCVNGGKCYHVSACLCGAPGRCERARLCPSGLYGIKDKRCPCHLENTHSCH 181
OY 380 PMSGECACRPGMSGLYCNETCSPGYGACQOICSCONG 418
DB 182 PMSGECACRPGMSGLYCNETCSPGYGACQOICSCONG 220

```

```

OY 200 GATCCHVTGECRCPPGYTGAFCECLCPGKHGPCEQORPCPCONGCVCHVTGECSPSGM 259
DB 2 GATCCHVTGECRCPPGYTGAFCECLCPGKHGPCEQORPCPCONGCVCHVTGECSPSGM 61
OY 260 MCTVCGOPCEPRFGKNCSEOCCHNGTCDATGQCCHCSPGYTGRCODECPVGTGYVL 319
DB 62 MCTVCGOPCEPRFGKNCSEOCCHNGTCDATGQCCHCSPGYTGRCODECPVGTGYVL 121
OY 320 CAETCOCVNGGKCYHVSACLCGAPGRCERARLCPSGLYGIKDKRCPCHLENTHSCH 379
DB 122 CAETRCVNGGKCYHVSACLCGAPGRCERARLCPSGLYGIKDKRCPCHLENTHSCH 181
OY 380 PMSGECACRPGMSGLYCNETCSPGYGACQOICSCONG 418
DB 182 PMSGECACRPGMSGLYCNETCSPGYGACQOICSCONG 220

```

OC Rhabditiidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Harris B.R.;
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99069613; PubMed=9851916;
 RA none;
 RT "Genome sequence of the nematode C.elegans: A platform for
 RT investigating biology."
 RL Science 282:2012-2018(1998).
 DR EMBL: AL032657; CAD27614.1;
 SQ SEQUENCE 1070 AA; 114180 MW; 75254D0DD5643A5 CRC64.

Query Match 35.7%; Score 1284.5; DB 5; Length 1070;
 Best Local Similarity 34.2%; Pred. No. 6.5e-116;
 Matches 246; Conservative 77; Mismatches 221; Indels 175; Gaps 20;

OY 21 GASPUNLEDPVCSHMSYSTVOESYPHPDQIYTT-----SCDILMMFCTRRH 73
 DB 35 GTTTP--QGDHVCT-----VKTIVDDY--ELKVIHTVYVNDTQCLNPLTGFQC---- 80
 OY 74 VSYRTAYRHGKTYRRK-----SOCCPFYSEGE-MCVPHCADKCVHGRCIAPNTC 124
 DB 81 ----TVERKGKASYQRLVKKEKYKCCDGYQTKHFLCDNPNPCKKCKIEPGKC 136
 OY 125 OCEPWGCGTNCSSACDGDHMGPHCTSRQCKNGALCNPTTGACHCACAGFRGRCE----- 179
 DB 137 ECDPQYGGKYYCASSCSVGTWGLGCKSCDCENGANCDELGTCTCTSGFQGERCKPCPD 196
 OY 180 -----DRCEGTGYNDCHQRCOCONGAT 202
 DB 197 NKMGPNVCKSCPCONGCKNKGKCVCSDMGGEFCCLKCEGKGAECFKCNCONGAT 256
 OY 203 CDHVTGECRCPPGYTGAFCEDLCPGKKGPOCEORPCONGVCHHVTGECSCPSGMWGT 262
 DB 257 CDTNMGKCKICKSGYHGALCENECISVGFSGGCTGCDLNNQNCSSSECKC-IGTWGK 315
 OY 263 VCGQCPBGRGRKNSQEQC-----HNGCTCDATGQCCHCSPTITGERCD-ECPVGT 315
 DB 316 HCDIGCSKRGFLQCKQKQNTCPGLEFSDSNASCDAKTCQCESEGYKPKCDERKCDLQ 375
 OY 316 YGVLCATCCQY--NGKCYHVSAGALCEAGFAGEGRCARLCPGLYIGCKDKRCPCHLE 373
 DB 376 YGADCSKTCCTCYRENTLMCAPTGTGCRCKPGFYGNCEL-ATSKDSYGNCKEKAMCDWN 434
 OY 374 NTHSGHPMSGECACRPGMSGLYCNETGSPGFYGEACQOITCSC-QNGADCDSTYTGKCTGAP 432
 DB 435 HASECNPEHSGCYCKPGRFGKNCSEPCPLDFYGPNCAMHQCQCNQNGVGGDADGKCCQDR 494
 OY 433 GFKGIDCSTPCLGTGYGINSRCCGKNDVAVSPVDGSCITCAAGHGVDCSTRCPSGTWG 492
 DB 495 GWTGRCHEHCADTFEAGNCEKRCCKPGKIGCDPTGTGECOTAGAGNCDIGCEGSGYG 554
 OY 493 FGCNLTCCOLINGAGCNTLDGTCTCAPW-----RGEKCEL--PCOD----- 531
 DB 555 PGCKLHCKCVN-GKCDKREGETCQPGFPGSDCTTCKSKRTGSECELSCPDSASCQ 613
 OY 532 -----GTGLNCAERCD----- 543
 DB 614 TGKCLCPGTGKGVSCDQCDPRTFGLQCEVTYPSPCASTDPKNGVCLSCPPGSGIHCE 673
 OY 544 -----CSHAD--GCHPTTGCHCCLPMSGVHCDSDVCAGRMGPNSLPG 585
 DB 674 HNCPRAGSTGDCQYVCSADHGCDPTTGECICEPGYHGKTCSEKCPDKTYGICALDC 732

DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Y47H9C.4 protein (CED-1).
 GN Y47H9C.4 OR CED-1.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditiida; Rhabditidae;
 OC Rhabditiidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Harris B.;
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Ahnscough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkneen R.,
 RA Smaldon N., Smith A., Sonhammer E., Staden R., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.,
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans."
 RL Nature 368:32-38(1994).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21097720; PubMed=11163239;
 RA Zhou Z., Hartweg E., Horvitz H.R.;
 RT "CED-1 is a Transmembrane Receptor that Mediates Cell Corpse
 RT Engulfment in C. elegans."
 RL Cell 104:43-56(2001).
 DR EMBL: AL032657; CAA21739.1;
 DR EMBL: AF332568; AAG60061.1;
 DR HSSP: P05106; 1JY2.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR002049; Laminin_EGF.
 DR Pfam: PF00008; EGF_9.
 DR PRINTS: PR00011; EGF_LAMININ.
 DR SMART: SM00180; EGF_Lam; 6.
 DR SMART: SM00261; FU; 2.
 DR SMART: SM00001; EGF-like; 5.
 DR PROSITE: PS00022; EGF_1; UNKNOWN_15.
 DR PROSITE: PS01186; EGF_2; 11.
 KV EGF-like domain; Glycoprotein..
 SQ SEQUENCE 1111 AA; 118803 MW; A39F37AC008F9874 CRC64;

Query Match 35.7%; Score 1284.5; DB 5; Length 1111;
 Best Local Similarity 34.2%; Pred. No. 6.7e-116;
 Matches 246; Conservative 77; Mismatches 221; Indels 175; Gaps 20;

OY 21 GASPUNLEDPVCSHMSYSTVOESYPHPDQIYTT-----SCDILMMFCTRRH 73
 DB 35 GTTTP--QGDHVCT-----VKTIVDDY--ELKVIHTVYVNDTQCLNPLTGFQC---- 80
 OY 74 VSYRTAYRHGKTYRRK-----SOCCPFYSEGE-MCVPHCADKCVHGRCIAPNTC 124
 DB 81 ----TVERKGKASYQRLVKKEKYKCCDGYQTKHFLCDNPNPCKKCKIEPGKC 136
 OY 125 OCEPWGCGTNCSSACDGDHMGPHCTSRQCKNGALCNPTTGACHCACAGFRGRCE----- 179
 DB 137 ECDPQYGGKYYCASSCSVGTWGLGCKSCDCENGANCDELGTCTCTSGFQGERCKPCPD 196
 OY 180 -----DRCEGTGYNDCHQRCOCONGAT 202
 DB 197 NKMGPNVCKSCPCONGCKNKGKCVCSDMGGEFCCLKCEGKGAECFKCNCONGAT 256
 OY 203 CDHVTGECRCPPGYTGAFCEDLCPGKKGPOCEORPCONGVCHHVTGECSCPSGMWGT 262

```

DB 257 CDTNKCJCKSCSYHNALECNESVGFEGSGCTQKCDCLNNONCDSSSGCKC-IGWTGK 315
263 VCGOPCEGRFGRKNCSEOC-----HNGCTDAATGOCHSPGYTGECOD-ECPVGT 315
DB 316 HCICIGSRGFRGLOCKONCTCPGLERSDSNASDAACTGQCQSGTKGPKCDERKDAEQ 315
316 YGVLCATCCOV--NGKCYHVSAGLCEAGFAGEARLCEPGLYGLIKDRCPCHLE 373
DB 376 YGADSKCTCTVHENTLMCAPNTPGCRCKRPGYDNCCL-AGSKDSYGPNCERQANCDM 434
374 NTHSCHPMSEACAKRPGWSLLYCNENCSPGFYEGACQITSC-QNGADCDVTGKCTCAP 432
DB 435 HASECNPETGSCYCKRGRGKNCSEPCPLDFYPNCAHOCQCNQRGVCGDAGDKCQCDR 494
433 GFGICDSTFPCPLGTGNGINCSRCCKNDAYCSPYDSCTCKAGHGVDSIRCPSTGW 492
DB 495 GMTGHRCEHCPCADPTTGANCERKCKPKGIGCPITGECTCPAGLOGANCDICPESSTG 554
493 FGCNLTCCCLNGACNTLDGCTCTCAPGW-----RGEKCEL--PCOD----- 531
DB 555 PGCKLMCKCVN-GKCKMETGECTCPGFFGSDSTCSKRYEGESCELSGCDASCSQ 613
532 -----GTGLNCAERCD----- 543
DB 614 TGKCLCPLGTGKVSQCDKCDPNTFGLCOETVTPSPCASTDPKNVGLSCPPSSGJHC 673
544 -----CSHAD--GCHPTGHCRLPGMSGVHCDSVCAEGHMGNCSLPC 585
DB 674 HNCPASGYDGCQOVCSADHGCDDPTTGECICEPGHGTSEKCPDKYTGICALDC 732

RESULT 13
075095 PRELIMINARY: PRT: 1246 AA.
AC 075095:
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-JUN-2002 (TREMBLrel. 21, last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)
DE MEGF6 (Fragment).
GN MEGF6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRIN;
RA MEDLINE=98360089; PubMed=9693030;
RA Nakayama M., Nakajima D., Nagase T., Nomura N., Seki N., Ohara O.;
RT "Identification of high-molecular-weight proteins with multiple EGF-
RT like motifs by motif-trap screening."
RL Genomics 51:27-34(1998).
DR EMBL: AB011539; BAA32467.2; -.
FT NON_TER 1
SO SEQUENCE 1246 AA; 130304 MW; BD8E70ADF6A2CB29 CRC64;

Query Match 35.6%; Score 1282.5; DB 4; Length 1246;
Best Local Similarity 41.8%; Pred. No. 1.2e-115;
Matches 216; Conservative 58; Mismatches 214; Indels 29; Gaps 7;

```

```

QY 258 GMMGTGCPPCPEGRGRKNCSEOCCHNGSTCDATGCHSPGYTGECODECPVYTG 317
DB 849 GMTGTCEHACPAFFGLDORSAACNCTAGAACAVANGSCLCPAGRGPRCAEKCLPRDR 908
QY 318 VLCAETCCYNGKCYHVSAGLCEAGFAGEARLCEPGLYGLIKDRCPCHLENTIS 377
DB 909 AGCRHSGGCLNGGLCDPHHGRCLCPAGWTGDKQSP-CLRGWGEACAPRCSC--PPAA 965
QY 378 CHPMSEACAKRPGWSLLYCNENCSPGFYEGACQITSCQNGADCDVTGKCTCAPFGKI 437
DB 966 CHHTACNCRCPFGFTSGCGCCPPRGYRPGCQDLGCLNGSSCDATGACRCPTEFLGT 1025
QY 438 DCSTPCPLGTGNGINCSRCCKNDAYCSPYDSCTCKAGHGVDSIRCPSTGWGCM 497
DB 1026 DCNLTCPGFRGPNCTHVCQGGGACADDPVTGCTCLPPRAGVGRGRCQPNRFGVCEH 1085
QY 498 TCQCLNGACNTLDGCTCTCAPGRGKCELPCDDGTGTLNCAERCCSHADGCHPTTGC 557
DB 1086 TCSGRNGGLCHASNGSCSGIGMTGHRHCELACPPRGYGAACHLECSHNNSTCEPATGTC 1145
QY 558 RCLPGW-----SGVHODSVCAGRMGWPNCSLPC 585
DB 1146 RCGPFGYGAACERHRSAGATCNLDCCRQGRFPSTLHC 1182

```

```

RESULT 14
09W0A0 PRELIMINARY: PRT: 434 AA.
AC 09W0A0:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE CG18172 protein.
GN CG18172.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RA MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Chame M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abail J.F., Adayanil A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Burris K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz G.H., Ke Z., Kennen J.A., Kethum K.A.,
RA Foster C., Garg P., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman J.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennen J.A., Kethum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mervulov G., Milshina N.V., Mobarry C., Morris J., Mostrel A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden Klamas I., Simpson M., Skupski M.P., Smith T.,
RA Spler E., Spradling A.C., Stapleton M., Strong R., Sun E.,

```



```

Db 161 PCPPGFYGMCKERCPEILHGKNSCDHITGIELCRTGYIGLTCHEPCPAGLYGPCKIKC 220
QY 239 PCQNGVYCHHYTGECSGSPGSMGTVCGOPCEGRFGKNSQECQCHNGTCDATGQCHC 298
    |::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 221 NCEHGGECNNHTGOCCLPMTGSCNESCPTDITYGOGCAQRCRCVHHKCYCKRADGMCIC 280
QY 299 SPGYTGERCODECPVGTGYVLCALCTCQVNGKCYHVSAGCLCEAGFAGERCEARLCPEG 358
    |::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 281 ETGWSGTRCDEVCEPGEFYGEHCHMNTACPSANFQCHAHGCVCRSGYTGDMCDELIASQR 340
QY 359 LYG-----IKCDKRCPCHELNTHSCHPMSCGE 384
    |::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 341 IADQSENSSRASVALTLVLTLEFACILFAVFITYRRRVSNLKTETIAHVHTHDTNP---- 396
QY 385 CACKPGWSGLYCNETCSPGFYG-EACQOICSCQNGADDCDSVTKRCTCAPGFKGIDCSTPC 443
    |::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 397 ----PSMP---NNFNPNPYGMAETRLPNNMRSKMNNPDQSTWSTDY-GDDCNASG 448
QY 444 PLGTGYINGCSSRCGCKN 460
    |::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 449 RVGSYSINYNHDLTLTKN 465

```

Search completed: May 9, 2003, 11:57:14
 Job time : 40.9328 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 9, 2003, 11:39:55 ; Search time 64.0672 Seconds

(without alignments)
3666.364 Million cell updates/sec

Title: US-10-092-390-2

Perfect score: 6744
Sequence: 1 MVSLNSCLSFICLLCHMT.....SSPKQDSGSSSSSSSSSE 1140

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: SPREMEL_21:*
2: sp.archaea:*
3: sp.bacteria:*
4: sp.fungi:*
5: sp.human:*
6: sp.invertebrate:*
7: sp.mammal:*
8: sp.mhc:*
9: sp.organelle:*
10: sp.phage:*
11: sp.plant:*
12: sp.podent:*
13: sp.virus:*
14: sp.vertibrate:*
15: sp.unclassified:*
16: sp.virus:*
17: sp.bacteriap:*
17: sp.archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6744	100.0	1140	4	Q96K77 homo sapien
2	3769	55.9	969	4	Q96K66 homo sapien
3	3468	51.4	567	4	Q8WUL3
4	2668	39.6	1034	11	Q8VHL7
5	2667	39.5	1034	11	Q8VHL7
6	2292.5	34.0	747	11	Q8VHL7
7	1958	29.0	1574	11	Q8VHL7
8	1931	28.6	1664	5	Q9TVQ2
9	1823.5	27.0	1070	5	Q8T3A7
10	1805.5	26.8	1111	5	Q9XWD6
11	1804.5	26.8	1246	4	Q75095
12	1804	26.7	1045	5	Q8T3A6
13	1309	19.4	220	11	Q63404
14	1252	18.6	434	5	Q9W0A0
15	1024.5	15.2	2447	13	Q13149
16	1006	14.9	2653	5	Q25253

17	988.5	14.7	2524	5	Q9PA5	Q9PA5 branchiosto
18	964.5	14.3	2531	5	Q16004	Q16004 lytechinus
19	920.5	13.6	4006	11	Q35452	Q35452 mus musculu
20	916.5	13.6	491	4	Q8TEK2	Q8TEK2 mus sapien
21	916.5	13.6	4288	4	Q9NPK9	Q9NPK9 homo sapien
22	915.5	13.6	2352	5	Q61240	Q61240 halocynthia
23	903.5	13.4	4114	11	Q54796	Q54796 mus musculu
24	881.5	13.1	4135	6	Q18977	Q18977 bos taurus
25	863	12.8	713	5	Q962M9	Q962M9 podocoryne
26	832.5	12.3	752	13	Q42374	Q42374 brachydantio
27	790.5	11.7	1214	13	Q90YD2	Q90YD2 xenopus lae
28	782	11.6	1254	13	Q9YHU2	Q9YHU2 brachydantio
29	782	11.6	1254	13	Q90Y56	Q90Y56 brachydantio
30	779	11.6	1193	13	Q90819	Q90819 gallus gall
31	765	11.3	1212	13	Q42347	Q42347 gallus gall
32	763.5	11.3	2146	5	Q9VC37	Q9VC37 drosophila
33	762	11.3	1216	13	Q90Y55	Q90Y55 brachydantio
34	748.5	11.1	2019	11	Q64706	Q64706 mus musculu
35	741	11.0	594	5	Q9W0A1	Q9W0A1 drosophila
36	740	11.0	594	5	Q9Y151	Q9Y151 drosophila
37	735.5	10.9	3695	4	Q8TDE8	Q8TDE8 homo sapien
38	723	10.7	3367	5	Q9XZC9	Q9XZC9 drosophila
39	720.5	10.7	1532	13	Q90994	Q90994 gallus gall
40	720.5	10.7	1714	13	Q90995	Q90995 gallus gall
41	711.5	10.6	2656	5	Q9GND3	Q9GND3 paracentrot
42	709.5	10.5	1810	13	Q90824	Q90824 gallus gall
43	707.5	10.5	1799	11	Q8RYO0	Q8RYO0 mus musculu
44	700.5	10.4	1792	13	Q57484	Q57484 gallus gall
45	696.5	10.3	3319	5	Q9VJ75	Q9VJ75 drosophila

ALIGNMENTS

RESULT 1	Q96K77	PRELIMINARY;	PRT; 1140 AA.
ID	Q96K77		
AC	Q96K77		
DT	01-DEC-2001 (TRMBREL_19, Created)		
DT	01-DEC-2001 (TRMBREL_19, Last sequence update)		
DT	01-JUN-2002 (TRMBREL_21, Last annotation update)		
DE	MEGF10 protein (K1A1780).		
GN	MEGF10.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=HIPPOCAMPUS; PubMed=11347906;		
RX	MEDLINE=21245130; PubMed=11347906;		
RA	Nagase T., Nakayama M., Nakajima D., Kikuno R., Ohara O.;		
RT	"Prediction of the coding sequences of unidentified human genes. XX.		
RT	The complete sequences of 100 new cDNA clones from brain which code		
RT	for large proteins in vitro."		
RL	DNA Res. 8:85-95(2001).		
DR	EMBL: AB058676; BAB47409.1; -		
DR	InterPro: IPR000561; EGF-like.		
DR	Pfam: PF00008; EGF_15.		
DR	PROSITE: PS00022; EGF_1; UNKNOWN_17.		
DR	PROSITE: PS01186; EGF_2; UNKNOWN_17.		
SQ	SEQUENCE 1140 AA; 122204 MW; 4582FA239423895A CRC64;		
Query Match	100.0%; Score 6744; DB 4; Length 1140;		
Best Local Similarity	100.0%; Pred. No. 0;		
Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
QY	1	MVSLNSCLSFICLLCHMTGTAAPLNEDPNVCSHESVSYVQESYPPFDIYTSC	60
DB	1	MVSLNSCLSFICLLCHMTGTAAPLNEDPNVCSHESVSYVQESYPPFDIYTSC	60
QY	61	TDILNFKCTRRHVSRYTAVRHGKTYRRKSQCCPFYEGSCVPHCADKCHGRGIA	120

```

Db 61 TDILNFKCTRRHVSRYTRAYRHGEKTMRYRKSOCCPFYEGSEGMVCPHCAKRCVHGRGRIA 120
Oy 121 PNTCCPEPMGNTGNCSSACDGDHMGPHCTSRCCCKNGALCNPTTGACHCAAGFRGRCD 180
Db 121 PNTCCPEPMGNTGNCSSACDGDHMGPHCTSRCCCKNGALCNPTTGACHCAAGFRGRCD 180
Oy 181 REBQGTGNDCHORCOQONGATCDHVTGECRCRPGYTGAFCEDLCPPGKHGPOCEQRCP 240
Db 181 REBQGTGNDCHORCOQONGATCDHVTGECRCRPGYTGAFCEDLCPPGKHGPOCEQRCP 240
Oy 241 ONGGVCNHTGECSCPSGSMGTVCCGPCEPGRGRKNCSECCCHNGCTCDAAATGQCHCSP 300
Db 241 ONGGVCNHTGECSCPSGSMGTVCCGPCEPGRGRKNCSECCCHNGCTCDAAATGQCHCSP 300
Oy 301 GYTERCODECEPVGTGVLCAETCCQVNGKCYHVSAGALCEAGFAGERCLEARLCEPEGLY 360
Db 301 GYTERCODECEPVGTGVLCAETCCQVNGKCYHVSAGALCEAGFAGERCLEARLCEPEGLY 360
Oy 361 GIKCDKRCPCHEBNTSHCHPMSEGCACRPMGSLYCNETCSPGFYGEACQOITCSCNGAD 420
Db 361 GIKCDKRCPCHEBNTSHCHPMSEGCACRPMGSLYCNETCSPGFYGEACQOITCSCNGAD 420
Oy 421 CSYTGKTCAPGFKGIDSTPCPLGTGYNCSRCGCKNDVCSVPDSCCTCKAGMHGV 480
Db 421 CSYTGKTCAPGFKGIDSTPCPLGTGYNCSRCGCKNDVCSVPDSCCTCKAGMHGV 480
Oy 481 DCSIRCPSTGWFSGCNLTCCCLNGAGCNTLDGTCTCAPGMRGECBLPCODGTGYLNCAB 540
Db 481 DCSIRCPSTGWFSGCNLTCCCLNGAGCNTLDGTCTCAPGMRGECBLPCODGTGYLNCAB 540
Oy 541 RCDSCNADCHPTTGRCRLPFGSHGYHCDYVCAEGRMGNCSPCTCKNGASCPDGTIC 600
Db 541 RCDSCNADCHPTTGRCRLPFGSHGYHCDYVCAEGRMGNCSPCTCKNGASCPDGTIC 600
Oy 601 ECAPGRGTTCORICSPGEGYHNRCSQTCPCYHSSGPRCHITGLCCLPFTGALCNEYC 660
Db 601 ECAPGRGTTCORICSPGEGYHNRCSQTCPCYHSSGPRCHITGLCCLPFTGALCNEYC 660
Oy 661 PSGRREGKNCAGICTCTNNCTNPIDRSQCQYPMGIGSDCSQPCPRAHMGPNCTIHTCNCHN 720
Db 661 PSGRREGKNCAGICTCTNNCTNPIDRSQCQYPMGIGSDCSQPCPRAHMGPNCTIHTCNCHN 720
Oy 721 GAFCSAYDECKCTPGMTLCTGTCORPLGFYGDALITCQONGACDHIISGCTCTRTGF 780
Db 721 GAFCSAYDECKCTPGMTLCTGTCORPLGFYGDALITCQONGACDHIISGCTCTRTGF 780
Oy 781 MGRHCEQKCPSTGYGCGHQICDCLNNSTCDHITGTCYCSPGMKGARCDQAGYIYGNLN 840
Db 781 MGRHCEQKCPSTGYGCGHQICDCLNNSTCDHITGTCYCSPGMKGARCDQAGYIYGNLN 840
Oy 841 SLSTRSTALPADSYOIGAIAGIILVLLVFLALFIIYRHOKGESSMPAVTYTPAMR 900
Db 841 SLSTRSTALPADSYOIGAIAGIILVLLVFLALFIIYRHOKGESSMPAVTYTPAMR 900
Oy 901 VYNADVTISGTLPHSNGNANSHYFTNPSTHTLTOCATSPHYNNRBMVTKSKNOLTV 960
Db 901 VYNADVTISGTLPHSNGNANSHYFTNPSTHTLTOCATSPHYNNRBMVTKSKNOLTV 960
Oy 961 NLKNVNPGRGPVDCGTGLTPADMKHGYLNLGAFGLDRSYWKSJLKLJGKNSEYNSN 1020
Db 961 NLKNVNPGRGPVDCGTGLTPADMKHGYLNLGAFGLDRSYWKSJLKLJGKNSEYNSN 1020
Oy 1021 CSLSSSENATIKDPVLIIPKSSBEGYEMKSPARBDSPYAEINNSTANRVYVEEPT 1080
Db 1021 CSLSSSENATIKDPVLIIPKSSBEGYEMKSPARBDSPYAEINNSTANRVYVEEPT 1080
Oy 1081 VSVYGVGFNNGRRLSDPDLPRKNSHIPCXYDLPRDSSSPKODSGSSSSSSSE 1140
Db 1081 VSVYGVGFNNGRRLSDPDLPRKNSHIPCXYDLPRDSSSPKODSGSSSSSSSE 1140

```

```

ID 096K6 PRELIMINARY; PRT; 969 AA.
AC 096K6;
DT 01-DEC-2001 (TRENBLREL, 19, Created)
DT 01-DEC-2001 (TRENBLREL, 19, Last sequence update)
DE 01-JUN-2002 (TRENBLREL, 21, Last annotation update)
DE ME6F11 protein (K1A1781).
GN Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euteria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRIN;
RX MEDLINE=21245130; PubMed=11347906;
RA Nagase T., Nakayama M., Nakajima D., Kikuno R., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XX.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro."
RL DNA Res. 8:85-95(2001).
DR EMBL; AB058677; BAB47410.1; -.
DR InterPro; IPR000561; EGF-like.
DR Interpro; IPR001281; Rleske.
DR Pfam; PF00008; EGF_15.
DR PROSITE; PS00022; EGF_1; UNKNOWN_17.
DR PROSITE; PS01186; EGF_2; UNKNOWN_17.
DR PROSITE; PS00200; RIESKE_2; UNKNOWN_1.
SQ SEQUENCE 969 AA; 101600 MW; 56DDJFEFL39C8209 CRC64;

Query Match 55.9%; Score 3769; DB 4; Length 969;
Best Local Similarity 58.6%; Pred. No. 1,2e-296;
Matches 600; Conservative 126; Mismatches 208; Indels 90; Gaps 4;

Oy 109 CADKCYHGRICAPNTCCPEPMGNTGNCSSACDGDHMGPHCTSRCCCKNGALCNPTTGACH 168
Db 109 CADKCYHGRICAPNTCCPEPMGNTGNCSSACDGDHMGPHCTSRCCCKNGALCNPTTGACH 168
Oy 28 CTEBCHNGVCSBPDTCHCERGMGRPCSSGCSDDHMGPHCTSRCCCKNGALCNPTTGACH 87
Db 28 CTEBCHNGVCSBPDTCHCERGMGRPCSSGCSDDHMGPHCTSRCCCKNGALCNPTTGACH 87
Oy 169 CAAGFRGWRCEDECEGTGNDCHORCOQONGATCDHVTGECRCRPGYTGAFCEDLCPPG 228
Db 169 CAAGFRGWRCEDECEGTGNDCHORCOQONGATCDHVTGECRCRPGYTGAFCEDLCPPG 228
Oy 88 CAAGFRGWRCEDECEGTGNDCHORCOQONGATCDHVTGECRCRPGYTGAFCEDLCPPG 147
Db 88 CAAGFRGWRCEDECEGTGNDCHORCOQONGATCDHVTGECRCRPGYTGAFCEDLCPPG 147
Oy 229 KHGPOCEQRCPSTGYGCGHQICDCLNNSTCDHITGTCYCSPGMKGARCDQAGYIYGNLN 288
Db 229 KHGPOCEQRCPSTGYGCGHQICDCLNNSTCDHITGTCYCSPGMKGARCDQAGYIYGNLN 288
Oy 148 SHGACELRCPCQNGGTCNHTGECSCPRGMGTGAVADCPRTGFCQNSQDCPHNGQ 207
Db 148 SHGACELRCPCQNGGTCNHTGECSCPRGMGTGAVADCPRTGFCQNSQDCPHNGQ 207
Oy 289 CDAATGCHCSPGYTERCODECEPVGTGVLCAETCCQVNGKCYHVSAGALCEAGFAGERCLEARL 348
Db 289 CDAATGCHCSPGYTERCODECEPVGTGVLCAETCCQVNGKCYHVSAGALCEAGFAGERCLEARL 348
Oy 208 CDHVTGCGCHTAGYMKDRQOEBPFGSPQSCDCHNGGQCSFTTGACDEPQYKRP 267
Db 208 CDHVTGCGCHTAGYMKDRQOEBPFGSPQSCDCHNGGQCSFTTGACDEPQYKRP 267
Oy 349 KCEARLCPEGLYGIKCDKRCPCHEBNTSHCHPMSEGCACRPMGSLYCNETCSPGFYGEA 408
Db 349 KCEARLCPEGLYGIKCDKRCPCHEBNTSHCHPMSEGCACRPMGSLYCNETCSPGFYGEA 408
Oy 268 RQERLCPEGLYGIKCDKRCPCHEBNTSHCHPMSEGCACRPMGSLYCNETCSPGFYGEA 327
Db 268 RQERLCPEGLYGIKCDKRCPCHEBNTSHCHPMSEGCACRPMGSLYCNETCSPGFYGEA 327
Oy 409 COOICSCONGADSDVTGCTCAPGFKGIDSTPCPLGTGYNCSRCGCKNDVCSVPD 468
Db 409 COOICSCONGADSDVTGCTCAPGFKGIDSTPCPLGTGYNCSRCGCKNDVCSVPD 468
Oy 328 COLPCTCQNGADSHSTIGCTCAPGFMGEYCAVSCAGTYGPNCSICSCNNGTCSVPD 387
Db 328 COLPCTCQNGADSHSTIGCTCAPGFMGEYCAVSCAGTYGPNCSICSCNNGTCSVPD 387
Oy 469 GSCTCAKAGHGYDCSIRCPSTGWFSGCNLTCCCLNGAGCNTLDGTCTCAPGMRGECBLPC 528
Db 469 GSCTCAKAGHGYDCSIRCPSTGWFSGCNLTCCCLNGAGCNTLDGTCTCAPGMRGECBLPC 528
Oy 388 GSCTCAKAGHGYDCSIRCPSTGWFSGCNLTCCCLNGAGCNTLDGTCTCAPGMRGECBLPC 447
Db 388 GSCTCAKAGHGYDCSIRCPSTGWFSGCNLTCCCLNGAGCNTLDGTCTCAPGMRGECBLPC 447
Oy 529 CODGTGGLNCAERCDCHADGCHPTTGRCRLPFGSHGYHCDYVCAEGRMGNCSPCTCKNGAS 588
Db 529 CODGTGGLNCAERCDCHADGCHPTTGRCRLPFGSHGYHCDYVCAEGRMGNCSPCTCKNGAS 588
Oy 448 CPDGTGGLNCAERCDCHADGCHPTTGRCRLPFGSHGYHCDYVCAEGRMGNCSPCTCKNGAS 507
Db 448 CPDGTGGLNCAERCDCHADGCHPTTGRCRLPFGSHGYHCDYVCAEGRMGNCSPCTCKNGAS 507
Oy 589 NGASCPDGDICECAPGFGTTCORICSPGEGYHNRCSQTCPCYHSSGPRCHITGLCCLPFTGAL 648
Db 589 NGASCPDGDICECAPGFGTTCORICSPGEGYHNRCSQTCPCYHSSGPRCHITGLCCLPFTGAL 648
Oy 508 NGASCPDGDICECAPGFGTTCORICSPGEGYHNRCSQTCPCYHSSGPRCHITGLCCLPFTGAL 567
Db 508 NGASCPDGDICECAPGFGTTCORICSPGEGYHNRCSQTCPCYHSSGPRCHITGLCCLPFTGAL 567
Oy 649 PFTGALCNEYCPSGRFRGNCAGICTCTNNCTNPIDRSQCQYPMGIGSDCSQPCPRAHMG 708
Db 649 PFTGALCNEYCPSGRFRGNCAGICTCTNNCTNPIDRSQCQYPMGIGSDCSQPCPRAHMG 708
Oy 568 PFTGALCNEYCPSGRFRGNCAGICTCTNNCTNPIDRSQCQYPMGIGSDCSQPCPRAHMG 627
Db 568 PFTGALCNEYCPSGRFRGNCAGICTCTNNCTNPIDRSQCQYPMGIGSDCSQPCPRAHMG 627

```



```

QY 709 GNCITTCNCHNGAFCSAYDGECKTPEWTGLYCTORCPGLGYGKCALICCONGACDD 768
DB 628 GRACFIACCHNGASCASADGACHCTPGMTGLFCTYRCFAAFGADCGVCCQONASCD 687
QY 769 HISSGCTCTRTGEMGRHCEOKCPSTGYGCRQICDCLNNSTCDHTTCTCSPGKGCAR 828
DB 688 HISSGCTCTRTGEMGRHCEOKCPSTGYGCRQICDCLNNSTCDHTTCTCSPGKGCAR 747
QY 829 DDAGVIVVNLSTSTSTALPADSVOIGAIIGIIVLVYFLALFIIYHKKOKGKS 888
DB 748 DDA-ALMELNLPYTKISALAEKRSVGAIVGIMLLEFIVVLGLFAMHRRROEKER 806
QY 889 SM-PAVYTPAMRVVADITISGTLPHSNGNANSHYFNPSTHTLTCATSPHYANRDR 947
DB 807 DLAPRVSTYPPAMRMTSTDVSL----- 828
QY 948 MIVTSKNNQLVNLKNVPGKRGVGDCTGLPADMKHGYLNLGAFGLDRS---YM 1003
DB 829 -----GACGMDRRQNTYIM 842
QY 1004 GKSLDGLKNSSEYNSNCSLSSSENPYATIKDPVILPKSBCGYEMKSPARSDPYAE 1063
DB 843 DKGFDDYKESVCSSTCSLNSSENPYATIKDPVILPKSBCGYEMKSPVHMGSPTD 902
QY 1064 INNSTSANRVYEPTVSVGVFNSNGRLSDPYDLPKSHIPCHYDLPLVRDSSSEP 1123
DB 903 VSLSTSNKNIEVEPTVSVGVFNSNGRLSDPYDLPKSHIPCHYDLPLVRDSSSEP 962
QY 1124 KOED 1127
DB 963 PSQD 966

```

RESULT 3

```

Q8WUL3 PRELIMINARY: PRT; 567 AA.
AC Q8WUL3;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Similar to MEGF10 protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=MUSCLE;
RA Strausberg R.;
RL Submitted (DEC-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL: BC020198; AAL20198.1; -
DR InterPro: IPR00561; EGF-like.
DR Pfam: PF002049; Laminin_EGF.
DR PRINTS: PR00011; EGF_LAMININ.
DR SMART: SM00181; EGF_11.
DR SMART: SM00180; EGF_Lam; 9.
DR PROSITE: PS00022; EGF_1; UNKNOWN_10.
DR PROSITE: PS01186; EGF_2; UNKNOWN_10.
SQ SEQUENCE 567 AA; 60797 MW; CEF2FB8CDB7CF627 CRC64;

```

Query Match 51.4%; Score 3468; DB 4; Length 567;

Best Local Similarity 99.8%; Pred. No. 1.5e-272; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MVISNLSCLSTICLLCHMIGTASPLNEDPNVCSHMESYVTVQESYPHPFDQIYITSC 60
DB 1 MVISNLSCLSTICLLCHMIGTASPLNEDPNVCSHMESYVTVQESYPHPFDQIYITSC 60
QY 61 TDLIMFKTRHRSVYRTAYRGEKTYRRKRSQCCPGFESEGMCPVPCADKCVHGRGIA 120
DB 61 TDLIMFKTRHRSVYRTAYRGEKTYRRKRSQCCPGFESEGMCPVPCADKCVHGRGIA 120

```

```

QY 121 PNTCOEPEMGWGTGNCSSACDGDHMGPHCTSRQCKNGALCNPTGACACAGRGMRCD 180
DB 121 PNTCOEPEMGWGTGNCSSACDGDHMGPHCTSRQCKNGALCNPTGACACAGRGMRCD 180
QY 181 RCEQGTGNDCHQROCOQNGATCDHYTGRCRPPGYTGAFCEBDCPPGKHGPOCEORPC 240
DB 181 RCEQGTGNDCHQROCOQNGATCDHYTGRCRPPGYTGAFCEBDCPPGKHGPOCEORPC 240
QY 241 QNGGVCCHVHTGSCSPSGMGVTCGQPCPEGRFRKNSQSECCHNGGTDAATGQCHSP 300
DB 241 QNGGVCCHVHTGSCSPSGMGVTCGQPCPEGRFRKNSQSECCHNGGTDAATGQCHSP 300
QY 301 GYTGRCODECPVGYGYVLCATCTCQVNGCKYHSGACLCBAGEGCEARLPEGLY 360
DB 301 GYTGRCODECPVGYGYVLCATCTCQVNGCKYHSGACLCBAGEGCEARLPEGLY 360
QY 361 GIKCKRCPCHELNTSHCHPMSGECACRPGMSGLYCNETGSPGYGEACQITSCONGAD 420
DB 361 GIKCKRCPCHELNTSHCHPMSGECACRPGMSGLYCNETGSPGYGEACQITSCONGAD 420
QY 421 CDSVTKCTCAGRCFGICDSTPCPLGTGYNCSRGCKNDVAVCSPTCKAGMHCV 480
DB 421 CDSVTKCTCAGRCFGICDSTPCPLGTGYNCSRGCKNDVAVCSPTCKAGMHCV 480
QY 481 DCSIRCPSTGWFCCNLFCQCLNGACNTLDGTCAPAPRBEKELPCQDGTGIANCAE 540
DB 481 DCSIRCPSTGWFCCNLFCQCLNGACNTLDGTCAPAPRBEKELPCQDGTGIANCAE 540
QY 541 RDCSHADGCHPTGHCRCCLPGMSGV 566
DB 541 RDCSHADGCHPTGHCRCCLPGMSGV 566

```

RESULT 4

```

Q8VHL7 PRELIMINARY: PRT; 1034 AA.
AC Q8VHL7;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Jedi protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL; TISSUE=TESTIS;
RA Kriytsov A.V., Zinovyeva M.V., Hendrikx J., Vliesser J.W.M.,
RA Belyavsky A.V.;
RL "Jedi is a novel DSL and EGF-like repeat motif-containing protein
expressed on non-differentiated hematopoietic cells."
RL Submitted (NOV-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF444274; AAL38571.1; -
DR InterPro: IPR00561; EGF-like.
DR Pfam: PF00008; EGF_9.
DR PRINTS: PR00011; EGF_LAMININ.
DR SMART: SM00181; EGF_15.
DR SMART: SM00180; EGF_Lam; 14.
DR PROSITE: PS00022; EGF_1; UNKNOWN_13.
DR PROSITE: PS01186; EGF_2; UNKNOWN_12.
SQ SEQUENCE 1034 AA; 110540 MW; 551AE5166AE0111 CRC64;

```

Query Match 39.6%; Score 2668; DB 11; Length 1034;

Best Local Similarity 42.7%; Pred. No. 2e-207; Mismatches 379; Indels 172; Gaps 16;

```

QY 14 LILCHWIGTASPLNEDPNVCSHMESYVTVQESYPHPFDQIYITSCDILNM---FCT 70
DB 7 LILALGLRLTGLTNSNDPNVCTFWSFTTTRKSHLRPESLLPABSCH--RPWEDPHPCA 64
QY 71 RHRVSRTAYRGEKTYRRKRSQCCPGFESEGMCPVPCADKCVHGRGIA 130

```

```

Db 65 QPTVYRTVYKQVYKMSRPLQCCRGYESRGACVPLQAEVHGRCVARNOCQCAPGW 124
QY 131 GGTNCSACDGDHNGPHTSRQCKNGALNPITGACIACAGFGMRCEDEGCTGND 190
125 RGGDCSSSCAGMGPQCDKFCCHGNSSCDPKSGACCPGSLQPPNCLQCPMGNHGA 184
QY 191 CHORCOONGATCDHVTGECRCPRGYTGAFCEDLCPGKHGRPOCEQRCPONGVCHHT 250
185 CQFPCQCY - GASCDPQDQACFCPPGRAGPSCNVPCSGTDFCPCRTYPCQNGVQDSQ 243
QY 251 GEGSCPGMGTGCGORPBERFGKNSQEQCHNGGCDATQCHRSPEYTERCODE 310
244 GSCCPRPMGVICSLPCBGFHNPCTQECRCHNGGLCDFTQCHCAPGYIDDRQOE 303
QY 311 CPVGTGVLCAETQCVNGKCYHVSGACLCCEAGFAGERBARCLPEGLYIKCDKRPC 370
304 CPVGRFGQDCAETQDCAPGARCFRANAGACLEHFTDRCTERICLPDRYLSQDEPCTC 363
QY 371 HLENTSCHPMSEGCACKPMGSLYCNETSPGFYGEACQOIGSCQNGADCDVTGKCTC 430
364 DPEHLSCHPMHGECSQCPGWAHGLHNESCPDTHGPGCOBHCICLHGGLADSGLCRC 423
QY 431 APFGKIDCSTPCPLGTGTCSSRCCKNDVCSPYDGSCTKAGMNGVDCSTRCSGT 490
424 APGTGPHCANLCPDPTGTCSSRCSENAIACSPIDGTCTCEGMRGNCVPCPLGT 483
QY 491 WGFNCNLTQCLNGACNTLDGTCTCARGMGEKCELPDODGTGYNLCABERDCSHADG 550
484 WGFNCNASCQCAHNDGVCSPGTGACTCTPMHGHACQLPCPRGQEGEASVCDCHSDG 543
QY 551 HPTGHCRLCPMGVHNDSCVACGRNGPNCSLCYCKNAGSCPPDDGICCAAGFGRTT 610
544 DPHVGCRCQAGMGTGCHLPCBGFEGANGCNTCTCKNGTCSENGNCCVAGFFGRPS 603
QY 611 CQRTCSGFGYGRCSQTCPOCVHSSGPHHTGTCLDCLPFTGALCNEVCSGRFNKA 670
604 CQRCPPRGGRKRCVQ----- 619
QY 671 GICTCTNN-GTCNPIDRSCQCPMGWIGSDSCQPCPRAHMGPNCTIHTCNHNGAFCSAYDG 729
620 --CCKNNHSSCHPSDGTCSCLAGMTGPDSEACRPGHMLKCSQCLQCHHGGTCHPDG 677
QY 730 ECKTPEMTGLYQTCORPLERFGKDCALIQCCQNGADCDHISGCTCTRTGFMHGECK 789
678 SCICTPEMTGPNCLGSCPRPFVGNCSQLCQCDLG----- 712
QY 790 PSGTYGCGRCQICDLNNTCDHTGTCTCSPGKRGARCDQAGVLIIGNLNTSLRSTAL 849
713 -----EMCHPQTGACVCPRGHSGADK-----MGSQESTIMPTS- 747
QY 850 PADSTOICAGIAGIILVVLFLALFIIYRHKQKRESSMPAVTYTPAMRVVNADYTIS 909
748 PVTNHSIGAAYIGLAVLTGLVALFALFGRQMKREHHLAVASTG-RLDSDYVMP 806
QY 910 GLTPHSGNANSHYFTNPBYHTLQCATSPHVNNRDMVTYKSNQNLVNLKNVPGK 969
807 DVSP-----SYSHYISNPBYHTLSQSPNRPBP-----KVPQSOLVYSSQAPRPS 853
QY 970 RGVGVDCGTLPADMKHAGYLINELGAFGLDRSY-----NGKS 1006
854 RAHGREHNTYLPADMKRRRPHRGASHLDRSTSCSTSHRNGPFCPHKGPISREGIGAS 913
QY 1007 LKDLGNSEYNSNCSLSSENPRYATIKDPPVILPKSSEGVEMKSPARBDSPYAEINN 1066
914 VMSL-----SSENPRYATIRDLPSLGRBEREGYEMKPPVSPRROSLH- 958
QY 1067 STANANRYVER-----TYSVVQGVFNNGRSLQDP-----YDLPRKSHIPECHY 1111
959 --LRDRQORQLOPORDSGTYEOPSPISHNESIGSTPRLPGLPQGYDSPKSHIHGYH 1016
QY 1112 DLLPVDRSSSPKQ 1125
1111 1111 1111

```

```

Db 1017 DLLPVHRPPSPSR 1030
RESULT 5
OBVTK5 PRELIMINARY; PRT; 1034 AA.
AC OBVTK5;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-JUN-2002 (Tremblrel. 20, Last sequence update)
DE 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE MGEF12.
GN MGEF12.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=LIVER;
RA Ivanova N.B., Lemischka I.R.;
RT "The global gene expression profiling of the hematopoietic stem
cell."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF440279; AL33583.1;
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR002049; Laminin_EGF.
DR Pfam: PF00008; EGF; 9.
DR PRINTS: PRO0011; EGF/LAMININ.
DR SMART: SM00181; EGF; 15.
DR SMART: SM00180; EGF_Lam; 14.
DR PROSITE: PS00022; EGF_1; UNKNOWN_13.
DR PROSITE: PS01186; EGF_2; UNKNOWN_12.
SQ SEQUENCE 1034 AA; 110580 MW; 714E5016848E4EAC CRC64;

Query Match 39.5%; Score 2667; DB 11; Length 1034;
Best Local Similarity 42.8%; Pred. No. 2.4e-207;
Matches 494; Conservative 110; Mismatches 378; Indels 172; Gaps 17;

QY 14 LLLCHMIGTASPLNLEDPNVCSSHWESYSVTVQESYPRPFQIYVCTDILNM---FKCT 70
7 LLLALGRLGLTINSNDPNVCTFWESTTTTKESHLLPFLRAESCH--RPMEDRPTCA 64
Db 71 RHRYSYTAIRHGEKTYRKRKSGCCPGFYSGEMCVHCDKCYNGKCIAPNTQCEPGW 130
65 QPTVYRTVYKQVYKMSRPLQCCRGYESRGACVPLQAEVHGRCVARNOCQCAPGW 124
QY 131 GGTNCSACDGDHNGPHTSRQCKNGALNPITGACIACAGFGMRCEDEGCTGND 190
125 RGGDCSSSCAGMGPQCDKFCCHGNSSCDPKSGACCPGSLQPPNCLQCPMGNHGA 184
QY 191 CHORCOONGATCDHVTGECRCPRGYTGAFCEDLCPGKHGRPOCEQRCPONGVCHHT 250
185 CQFPCQCY - GASCDPQDQACFCPPGRAGPSCNVPCSGTDFCPCRTYPCQNGVQDSQ 243
QY 251 GEGSCPGMGTGCGORPBERFGKNSQEQCHNGGCDATQCHRSPEYTERCODE 310
244 GSCCPRPMGVICSLPCBGFHNPCTQECRCHNGGLCDFTQCHCAPGYIDDRQOE 303
QY 311 CPVGTGVLCAETQCVNGKCYHVSGACLCCEAGFAGERBARCLPEGLYIKCDKRPC 370
304 CPVGRFGQDCAETQDCAPGARCFRANAGACLEHFTDRCTERICLPDRYLSQDEPCTC 363
QY 371 HLENTSCHPMSEGCACKPMGSLYCNETSPGFYGEACQOIGSCQNGADCDVTGKCTC 430
364 DPEHLSCHPMHGECSQCPGWAHGLHNESCPDTHGPGCOBHCICLHGGLADSGLCRC 423
QY 431 APFGKIDCSTPCPLGTGTCSSRCCKNDVCSPYDGSCTKAGMNGVDCSTRCSGT 490
424 APGTGPHCANLCPDPTGTCSSRCSENAIACSPIDGTCTCEGMRGNCVPCPLGT 483
QY 491 WGFNCNLTQCLNGACNTLDGTCTCARGMGEKCELPDODGTGYNLCABERDCSHADG 550
484 WGFNCNASCQCAHNDGVCSPGTGACTCTPMHGHACQLPCPRGQEGEASVCDCHSDG 543

```



```

Db      762  TTKCAGAGVGNLCDDPCFAGYGYDCEOKSCADVASPHKSKVCHHYTGCTCGLCKTGP 841
Oy      263  VCGQPCPEGRFGKNCSEOCCHNGSTCDAATGQCHSCSPGYGERCODECPVGTGVLCAE 322
Db      842  LCDQSCAANTYGPNCIAHCSGVNCAKCDSDSCCHCTCFGYATCSSEVCPTRGFGIDCMQ 901
Oy      323  TCOOANGKCYHVSGACCEAGFAGEBCEARLCPREGLYGINKDRKCPCHLEHTHSCHEMS 382
Db      902  LCKCNMGAICDTNSGSCCEAPGWSGKCD-KACAPGTGCKDCKDC-ADGMH-CDPSD 958
Oy      383  GEACKPGMSGLYCNETSPGFYGEACQOICSQNGADCDSDVTGKCTCAGPCK----- 435
Db      959  GEICIPPKKHKKHCKDETDGSLFGAGCKGICSCQNGATCDSDVTGSECEGRPKRKKCDP 1018
Oy      436  -----GIDCSTPCPLGTG 449
Db      1019  CPDGRFGEBCNAICDCTTNDTSMYNPVARCDHYTGECRCPCAGWTGPDCCQTCPLGRHG 1078
Oy      450  INCSRCGCKNDVAVSPVDGSCOTKAGMHGVDCSIRCPSGTMGFPCNLCOCLNGACNT 509
Db      1079  EGRHSCQCSNGASCDRYTGFCDPCSGFMGNKSECEPBGMLWSNCKMHCICMHGECNK 1138
Oy      510  LDGTCTCAPGMRGECCELPQODGTGTLNCAERDCSHADGCHPTTGHCRLCPGMSGVHCD 569
Db      1139  ENGDCEDIDGWTGSCFELCPGQGFGRNCAQKNCNGASCDRKTRGRCBCLPGMSGHCE 1198
Oy      570  SYCAGBRGPNCSLPCYCKNGASCSPDDGICECAGFGTGTQORICSPGFYGHRSQTPC 629
Db      1199  KSCVGHYGAKEETCEBENGALCDPIGHCSCOPGMRGKCKNRCPLGFGYGRHCSQSC- 1257
Oy      630  QCVHSSGPHHTTGLCDCLPFTGALCNEVCP----- 661
Db      1258  RCANKS-CDHISGCGCPKGTAGHSTELCPDGTIFGSCSCCKCCGGENSKMDAISGKF 1316
Oy      662  -----SGRFGKNACIGCTNNGCTNPIDRSQCCYPGMIIGSDCSCPBA 706
Db      1317  CKPGHSGSDCKSGCVQGRFGPDCNQLCSENGCVDSSTGSCVCPGYITGCEIACQSD 1376
Oy      707  HMRPCIRHCNCHNCAFSAHYDECKCPGWTGLYCTQRCPLGFGKCALICQONCAD 766
Db      1377  RGPFCERKICENGSTDRILGQRCPLGFTGTCNOYCPGGRGACKCKEKCANG-H 1435
Oy      767  CDHISGQCTGRTGFMGRHCEOKCPGTYGYGCRQICDCLNNSCTHITGYCSPGMKA 826
Db      1436  CNASSGCKCNLGTGFPSCDSCPSGKYGICLNLDCBCTGQARCDPYGCGCCCPGRGTS 1495
Oy      827  RC 828
Db      1496  RC 1497

```

```

RESULT 9
Q8T3A7 PRELIMINARY; PRT: 1070 AA.
AC Q8T3A7;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Y47H9C.4B. protein.
GN Y47H9C.4B.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida; Rhabdilitidae;
OC Rhabdilitidae; Peloderlineae; Caenorhabdilitis.
OX NCBI_Taxid=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Harris B.R.;
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;

```

```

RT "Genome sequence of the nematode C. elegans: A platform for
RT Investigating biology."
RL Science 282:2012-2018(1998).
DR EMBL; AL032657; CAD27614.1;
SQ SEQUENCE 1070 AA; 114180 MW; 75254DD0D5643A5 CRC64;

Query Match 27.0%; Score 1823.5; DB 5; Length 1070;
Best local similarity 32.4%; Pred. No. 5.5e-139;
Matches 362; Conservative 155; Mismatches 416; Indels 183; Gaps 34;

Oy 21 GTASPLNEDNVNCSHMSYSTVQESYPHFQDYTYT-----SCTDILNMFECTRHR 73
Db 35 GTTEP-----QSHVCH-----VKTIVDY--ELKVIYHYVNDTEQCLNPLRFGC----- 80
Oy 74 VSYRTAYHGEKTMTRK-----SQCCEFTESGE-MCVPHCADKCVHGRCTAPMTC 124
Db 81 -----TVERKQKASAYORLVRKKEKVKCCJBYOTKHFFCLPDCNPPCKKCKLEPKC 136
Oy 125 OCEPGMGSTNCSACDGDHMGPHCTSRCCCKNGALCNPLTGACHCAAGFRGRCEDRCEQ 184
Db 137 EDDPEYGGKYCASSCSVGTWGLGCKSCDCENGACNDBELGCTCTSGFQGERCKPCPD 196
Oy 185 GTYNDCHQRCQCONGATCDHYTGECRCPPGYTGAFCEDLCPGKHGFCQCEORCPQONG 244
Db 197 NKMGNVCYKSCPCQNGKCNK-EGKVCVSDGSGEFLCNKEBKFGABCKFEKCNQNGA 255
Oy 245 VCHAYTGECSCPSGMGTVCOPCEPGRFGKNCSEOCCHNGGTDAATGQCHSPGTG 304
Db 256 TCDNNTNGKICIKSGHGALENESVGFSGCTQCKCLLNQNDSSSGCKCI-GWTG 314
Oy 305 ERQDECPVGTGYVCAETCCV-----NGKCYHVSAGALICEAGFGERCEARLCPGEG 358
Db 315 KHCODIGSRGRGLQCKNCTCPLGLEFSDNASCDARTGQOCCEGYGPKDEKCAAE 374
Oy 359 LYGIKDCRCPCHLENTSCHPMGEGCACRPMGLYCNETCSPGTYEAC--QOICSCQ 416
Db 375 QYGAQCSKTCYVRENTLMCAAPNTGFCRCKRFGYDNCLELACKSKDSTPNCCKQMDN 434
Oy 417 NGADCDSTGTGCTCAPRGKIDCSNPCLGTGTGICSSRCCCKNAY-CSPVDGSCCTKA 475
Db 435 HASBCNPETGSCVCPGRTGKNCSEPCPLDTPGPCAQCOCNORGVCDGADGQCQDR 494
Oy 476 GWHGVDCSIRPSTGFWGFCMLTQCLNGACNLTDTGCTCAPGMRGECCELPQDGTG 535
Db 495 GWTGHRCHHCRPADFFGANCERKCKRGKIGCDPTTGECTCPAGIAGANCOIDGCESSYG 554
Oy 536 LNCAERDCSHADGCHPTTGHCRLPGWSGVHCDVCAEGRMPCSLPYCKNGASCSP 595
Db 555 PGCKLHCKCVNGK-CDKETGECTQPGFPGSDCSTYCKSKRYGSECELSPCSD-ASCSK 612
Oy 596 DDGICECAPRGRTTCQRTICSPGFYGHRCQTCPCVHSSGPHHI---TGIC-DCLGTF 651
Db 613 QTGKCLPLGTGKVSCHDCKDKNFTGFLQETV-----TPSCASTDRKNGCISCPGGS 667
Oy 652 TGLCNEVCPSGRFGKNCAGICTNNGCTNPIDRSQCCYPMGWSIDSCOPCPAHMGP 711
Db 668 SGIRHEHNCRAGSYDGOQVCSADGHCDBPTTBECICEBRYHKTSEKCPDCKYTG 727
Oy 712 CIHTC-NCNHAFCSAHYDECKCPGWTGLYCTQRCPLGFTGKCALICQONG-ADCDH 769
Db 728 CALDPCKASGSTCDHNGILICIPAGLEGALCTRCBSAGFMNGCRQVCRCTSEKQCA 787
Oy 770 ISGQCTGTGMRGRICEKCPGTYGYGCRQICDCLNNSCTHITGYCSPGMKA 827
Db 788 QTEGSCPAGFOGDRCDPCEDGTGPPCIRKCKQGTATSSCNVSGACHCHPEFTGEF 847
Oy 828 CDOAGVILVGNLNSLSTST-----ALP-----ADSYQIAGIAG----- 861
Db 848 C-----HALCPSTFGKCSKRCPPDGGCGDGECDAAIGCHVDMSCGKAKOE 896
Oy 862 -----IILVLVLF-----LIALFTYHKK-OKGESSMPAVTYTPAMRV 901
Db 897 FEALNGAGRTGLTWFFVLLIIVALLGGLIALF--YNNKYOKERKDPMPYVSF----- 948

```

QY 902 VNADYITSGTLPHSNGNANSHFTNPSYHTLQCATSPHYNNRDMRYTKSKNNOLFVN 961
 Db 949 -----HKAPNNDGREFONPLX-----SROSVFP-----DSDAFSSENNNGH----- 986
 QY 962 LKRVNPGKRGVPDCTGTLPADMK-----HGGYLNELGAFGLDINSYMK-----SLK 1008
 Db 987 -----GGPPNGLLTLEEBELENNKRIHG-----RSAGRGNNDYASLD 1023
 QY 1009 DLCKNSEYNSNCSLSSSENPYATIKDPYLIKSS 1044
 Db 1024 EVAGBESSSSASASARENPYADISSPDPYTONSA 1059
 RESULT 10
 ID 09XMD6 PRELIMINARY: PRT: 1111 AA.
 AC 09XMD6:
 DT 01-NOV-1999 (Tremblrel. 12, Created)
 DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Y47H9C.4 protein (CED-1).
 GN Y47H9C.4 OR CED-1.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OC NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Harris B.
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94150718; Pubmed=7906398;
 RA Wilson R., Alnsough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Lister N., Latreille P.,
 RA Lightning J., Lloyd C., Murray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
 RA Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Wellstock L., Wilkinson-Sproat J., Wohlman P.,
 RT 2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RL Nature 368:32-38(1994).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21097720; Pubmed=11163239;
 RA Zhou Z., Hartwig E., Horvitz H.R.,
 RT CED-1 is a Transmembrane Receptor that Mediates Cell Corpse
 Engulfment in C. elegans.
 RL Cell 104:43-56(2001).
 DR EMBL: AL032657; CAA21739.1;
 DR EMBL: AF332568; AAG60061.1;
 DR HSSP: P05106; JUV2.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR002049; Laminin-EGF.
 DR Pfam: PR00008; EGF_9.
 DR PRINTS: PR00011; EGF_LAMININ.
 DR SMART: SM00180; EGF_Lam; 6.
 DR SMART: SM00001; EGF_Like; 5.
 DR SMART: SM00261; FU; 2.
 DR PROSITE: PS00022; EGF_1; UNKNOWN_15.
 DR PROSITE: PS01186; EGF_2; 11.
 KW EGF-like domain; Glycoprotein.
 SQ SEQUENCE 1111 AA; 118803 MW; A39F374C008E9874 CRC64;
 Query Match 26.8%; Score 1805.5; DB 5; Length 1111;
 Best Local Similarity 31.8%; Pred. No. 1.7e-137;
 Matches 373; Conservative 162; Mismatches 423; Indels 215; Gaps 40;

QY 21 GTASPLNLEDPNVCASHMESYSTVOESYPAPEDOIYTT-----SCTILANFKCTRRH 73
 Db 35 GTEPR---GGDHCT-----VKTIYDDY---ELKVIHTYVYNDTECLNPLRTFGQC----- 80
 QY 74 VSTRATVIRRGKTYMRK-----SQCCPGFESGE--MCVPHCKDKVHGNCIAPNMC 124
 Db 81 ----TVEKRQOKKASYOQRLVKKEKYVQCCDDGYQTKRHFCLDPCNPPCKKGCICPGKC 136
 QY 125 QCEPBGWGTNCSACGDDHMGPHCTSCCOCKNALCNPLTGACHCAAGFRGWRGCEPCRG 184
 Db 137 ECDPPTGGKTYCASSGVGTWGLGCSKSDENANCDPELGTICTGSGFGEKCEKPCPD 196
 QY 185 GTYGNDCHORPCONCATDHYTBECCRCPPGYGAFCEBLCPPKRGPOCEORPCONG 244
 Db 197 NKMGPNCVASCPCQNGGKCK--EGKCVSGDWGEGECLNKEGKRGAGEKFCNCQNGA 255
 QY 245 VCHHTVTEGSCPSGSMGYVCGOPCPGRGRKNCSDQCCHNGSTCDAAATGQCHSPGYTG 304
 Db 256 TCDNTNGKCIKCKSGYHGALCENECVGFSGCTOKCDCLNNOGNCSSSGECKCI--GMTG 314
 QY 305 ERCQDCPVGTYGVLAETGOCV-----NGKCYHVSAGCLCEAGFAGEBRCARLCPSG 358
 Db 315 KNCDDICSRGRFGLQCKQNTCTPGLFESDSNASCAKGTGQCESGKGPCKDERKCAE 374
 QY 359 LVGINKDKRCPCHLENTHSCHPMSGECACKPGMSGLYCNETCSPGFYGENC--QOICSCQ 416
 Db 375 QYGADCSKTCTCVRENTLMCAPNTGRCRCRPGYGNCELACGKSDSYGRNCEKQANCDNN 434
 QY 417 NCADCSVTGKCTCAPGFGIGDCTSPCLCTGYNCSRGCKANDAV--CSPVDGCTCA 475
 Db 435 HASECNPETGSCVCKPGRGKNCSEPCPLDFYPNCAHOCQNRGVGCDGADGKCCQCR 494
 QY 476 GHWGVNCSIRCPSGTWGFGCNLTCCCLNGACATLTGCTCAGMWRGECCELPDQGTG 535
 Db 495 GMTGHRCEHHCPRADTGANCKRCKPKGICGPIYEGCTCPAGLQAGANDIGCPGSGTG 554
 QY 536 LNCABRCDSSHADGCHPTTGHCRCLPGWSGVHDSVCAEGRMGPNCSPCYCKNGASCP 595
 Db 555 PGCKLHKCVKNGK--CKETGECTCGQFPGSDSTCSKGYESELSPCSD--ASCSK 612
 QY 596 DSGICECAGFRKTTQORLCSPEFYGRSQCPCQVHSGPCNHL---TGLC--DLRPF 651
 Db 613 QTKCKCLPGLTKGVSDQCKDPMTFGLQETV----TPSPASTDPKNGVCLSPSPS 667
 QY 652 TGAICNEVPSGGRFKNCAIGICTNNGTGCPNIPDRSCOCYPGIGSDSOPCPAPMGPN 711
 Db 668 SGIHCHENCPAGSYGDCQOVCSADGHGCDPTTGECICEBPGHGTSEKCPDGKTYG 727
 QY 712 CHTTC--NCHNGAFCSAYDECKCTPGMTGLYCTQRCPLGFGKDCALICQONG--ADCDH 769
 Db 728 CALDCPKCASGSCDHIINGLICIPALLEGALCTRPCSAGFWGCGCQVCRCTSEYKQNA 787
 QY 770 ISGQCTCRGFMNRHDEQKCPSTYVYGRQIOLDCLNST---CDHTTGCTCYSPGKGR 827
 Db 788 QTEGECSPAGFQDRCDDKPEDYIPDXDKKCKCGTATSSCNRVSGACHCPGTGEF 847
 QY 828 CDQAGVLIIVGNLSLSRTST-----ALP-----ADSYOIGATAG----- 861
 Db 848 C-----HALCPSTFGLKCKECPKDCGCGIECDALIGCHYDQMSCKRAOE 896
 QY 862 -----IIIVLVVLF-----LALFTIYRRK--QKGRSSNPAYTYTPAMRY 901
 Db 897 FEALNGAGRSTGLTFWFLVLLVLCGLGLIALF--YRNKYQEKDPDMPTVSF----- 948
 QY 902 VNADYITSGTLPHSNGNANSHFTNPSYHTLQCATSPHYNNRDMRYTKSKNNOLFVN 961
 Db 949 -----HKAPNNDGREFONPLX-----SROSVFP-----DSDAFSSENNNGH----- 986
 QY 962 LKRVNPGKRGVPDCTGTLPADMKHGYLNELGAFGLDINSYMK-----GSKLKDGLKN-- 1013
 Db 987 -----GGPPN--GLLTLEEBELENNKRIHGSAAGRNNDY 1019

Qy	305	ERCODEPVGVYGLVLAETQCV-----NGRCYVYSGACLCAEAFAGERCARLCPEBG	358
Db	315	KHCDIGSGRGFGTGCQKQNTCTCPGLEPESDSNASCDKATGQCCQCESSYKPKPKDERKCDAE	374
Qy	359	LVGICDKRCRCPHLENTHSCHPMSGECAKPGMSGLYCNETSPGFGYBAC--QQICSCQ	416
Db	375	QYGDADCKTCTCAVENTLMLCAPMTGCRCKRPGFYGVNCELASKDSYGVNCEKAMCDMN	434
Qy	417	NGACDVSYTGKCTCAPBEFKRIDSTCPRLGTGYINSSSGCKKNDAV--CSPVDSCTCKA	475
Db	435	HASCSNPETGSCVCKPRTKKNCSERPDLFTYRPNCAHQCCQKQKGVGDGADGKCQCDR	494
Qy	476	GMHWGDCSIRCPSGTGWGFCNLTCOCLNGACNTLDGTCTCAPWNGEKELCPCDDGTGY	535
Db	495	GWTGHRREHNCPADTFGANCERKCRKCPKIGICDPRTIGECTCAPAGLQAGNCIDCGEBSYG	554
Qy	536	LNCAERDCCSHADCSHTTTHCHRCILRWMSVHDSVCAGBRGKPNKSLPYCKKNAGASCP	595
Db	555	PGCLFLKCKVNGK--CDKETEBCCTQPEFFSDSCTTSCGKKGESSELSCPSPD--ASQSK	612
Qy	652	TGALCNEVCSGRRGKAKCAICTCTNNNGTCNPIDNSCQCYPCWIGSDGSCPCPPAHMGPN	711
Db	668	SGICEHNCPGASGSDCCQOVCSACBHGCDPRTTGECICEPGRHGTCKSEKCPDCKYGY	727
Qy	712	CIHRC--NCHNAGFSAVDGCKCTPGMTGLCYQRCRPLGFGYDCLLIGCCONG--ADDDH	769
Db	728	CALDCPKCASGSTDHNLNGLICIPAGLEAGLCTRPCSAGWNGCQKQVCRCTSEYKQCA	787
Qy	770	ISGQCTRTGFMHGRHCQKCPSGTGYGCRQICDLNLSNT--CDHITGTCYCPSPMKGAR	827
Db	788	QTGCEGCPAGQGRCKRCPEDGYGPDCKIKKCKCGTATSSCNRRYSAGACHHPFTGBF	847
Qy	828	CDQAGVIYVGNLNLSTST-----ALP-----ADSYQIGAIAG-----	861
Db	848	C-----NALCESTFGLKCSKECPKDCGCGYEDDAIGCCHVDNOSCGKANOE	896
Qy	862	-----IIILVLYLF-----LALFIYIRRK--OKGESSNPATYTPAMKV	901
Db	897	FEALNGAGRSYGLWPFVLLIYALCGGLIALF--YRNKYQEKEXDPDPTVSF-----	948
Qy	902	VNADYITISGLPHSNGSNASHFTFNPSTYTLTQCATSPHYNNRDMATYTKSKNNQLYN	961
Db	949	-----HAPANNDEGRFQNPFLY---SRQSVFP---DSDAFSESSNGNHO---	986
Qy	962	LKNVNPCKRGVPGDCTTLPADKMHGGLYNELGAFILDRSYMGKSLDKGKXSEVSSNC	1021
Db	987	-----GGPPH--GLTLLEEEL-----ENKKIHGNSA	1011
Qy	1022	SLSSSENPATIKDPVLIKSS	1044
Db	1012	AGRGKNKPYADISSPDPVTONSA	1034
RESULT 13			
063404			
ID	063404	PRELIMINARY;	PRT; 220 AA.
AC	063404;		
DT	01-NOV-1996 (Tremblrel. 01, Created)		
DT	01-NOV-1996 (Tremblrel. 01, Last sequence update)		
DT	01-DEC-2001 (Tremblrel. 19, Last annotation update)		
DE	(clone REMA) ORF (Fragment).		
OS	Rattus norvegicus (Rat).		
OC	Eukaryota; Metazoa; Chordata; Granulata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
OX	NCBI_Taxid=101116;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=HOLZMAN; TISSUE=BRAIN;		
RC	MEDLINE=96235155; PubMed=8642059;		
TX			

Query Match	Best Local Similarity	19.4%	Score 1309;	DB 11;	Length 220;
Matches 209;	Conservative	4;	Mismatches 6;	Indels 0;	Gaps
QY 200	GATCDHVTGECRCPRGYTGAFCEDLCPPKHKHPDCEQRCPCQNGVYCHVTGSCSPSGM	259			
DB 2	GATCDHITGECRCSPGYTGAFCEDLCPPKHKHPDCEQRCPCQNGVYCHVTGSCSPSGM	61			
QY 260	MGTVGQRCPCRGKNCSDQSCQCHNGCTDDAANGCHCSPGYTGRCQDCEPCVGYGVY	319			
DB 62	MGTVGQRCPCRGKNCSDQSCQCHNGCTDDAANGCHCSPGYTGRCQDCEPCVGYGVY	121			
QY 320	CAETCQCVNGKRCCHVSGACICEAGFAGEKCEARLCPBGLYGIKCRKCPCHLENTHSCH	379			
DB 122	CAETCQCVNGKRCCHVSGACICEAGFAGEKCEARLCPBGLYGIKCRKCPCHLENTHSCH	181			
QY 380	PMSGECACKPMSGLYCHNCSPSPFYEGACQOICSCQNG	418			
DB 182	PMSGECACKPMSGLYCHNCSPSPFYEGACQOICSCQNG	220			
RESULT 14					
Q9W0A0	PRELIMINARY;	PRT;	434	AA.	
AC Q9W0A0					
DT 01-MAY-2000	(TrEMBLrel. 13, Created)				
DT 01-MAY-2000	(TrEMBLrel. 13, Last sequence update)				
DT 01-DEC-2001	(TrEMBLrel. 19, Last annotation update)				
DE CG18172	protein.				
GN CG18172					
OS Drosophila melanogaster (Fruit fly).					
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;					
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachyoptera; Muscomorpha;					
OC Ephydroidea; Drosophilidae; Drosophila.					
NCBI_TaxID=7227;					
RP	SEQUENCE FROM N.A.				
RC STRAIN=BERKELEY.					
RX MEDLINE=20196006; PubMed=10731132;					
RA Adams M.D., Celinkler S.E., Holt R.A., Evans C.A., Gocayne J.D.,					
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,					
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,					
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,					
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,					
RA Abell J.F., Agbayan C., Baxter E.G., Helt G., Nelson C.R., Milos G.L.G.,					
RA Baller R.M., Basu A., Bakendale J., Bayraktaroglu L., Beasley E.M.,					
RA Beeson K.Y., Beron P.V., Berman B.P., Bhandari D., Bolshakov S.,					
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,					
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,					
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,					
RA de Pallos B., Delcher A., Deng X., Mays A.D., Dew I., Dietz S.M.,					
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,					

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwam C.,
 RA Jalili M., Kalush F., Karpén G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Matzel B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Meruliov G., Mishina N.V., Modyarty C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacble J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Relneert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of Drosophila melanogaster."
 RL Science 287:2185-2195(2000).
 DR EMBL; AE003472; AAF47553.1; -.
 DR HSP; P02468; IKLO.
 DR Flybase; FBgn0035261; CG18172.
 DR InterPro; IPR001774; DSL.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR002049; Laminin_EGF.
 DR Pfam; PF00008; EGF; 8.
 DR PRINTS; PR00011; EGF_LAMININ.
 DR SMART; SM00051; DSL; 1.
 DR SMART; SM00180; EGF_lam; 3.
 DR SMART; SM00001; EGF_Like; 6.
 DR PROSITE; PS00022; EGF_1; 7.
 DR PROSITE; PS01186; EGF_2; 9.
 KW EGF-like domain; Glycoprotein.
 SO SEQUENCE 434 AA; 46417 MW; 26B5FD1E5C7970E4 CRC64;

Query Match 18.6%; Score 1252; DB 5; Length 434;
 Best local Similarity 45.6%; Pred. No. 3, 6e-93;
 Matches 190; Conservative 55; Mismatches 170; Indels 2; Gaps 1;

DB 151 RCGKRGALCNPTGTGCHCAAGRGRCRDRCGTGYGNDCHGRCQCCQNGATGCDHTGRC 210
 2 QCCCLLNNAVCPEPSSGDCBCKAGTGTGACADICPEGFGANCSKRCRCNGKCHHVSQEC 61
 OY 211 RCPGTYGAFCEDLCPGKHGHPQCEORPCQNGVCHVHTYEGSCSPSGMWTGCGOPCE 270
 DB 62 QCAFGTGLCDMRGCDGRHGAQCCQDDPCQNDGKQCPETGACMCNRPWTGVDCAKPCY 121
 OY 271 GRRGKNCQBCQCHNGCTDATGCHSBSGTGRCQDECPVGTGVLCAETGCGVNG 330
 DB 122 GSYGSPCOSCECYKCAPRHHTGQCECPYGRGRCFDECOQINTGYFNCSTDCANPA 181
 OY 331 KCYHVSAGACCEAGFAGEGCEARLPEGLYGIKDKRCPCHEHTSHSCHPMSECAKCR 390
 DB 182 MCDRANGTCTCNGWTGAKCAERICANKYGLDCNRTCECDMHTDLCHETENCCQSTG 241
 OY 391 WSLGLYNETCSPGFYEGACQOLSCONGADCVTKCTCAFGKIDESTPCLTGTGT 450
 DB 242 WSSAQCTRCPTFLRYGPNCLELNCNKAQKSPVNGTCLCAQWRBPTEESESEPGTFQ 301
 OY 451 NGSRRGCKNDVCSFYDQSGCTCKAGWGVDCIRPSTGWTGCGNLTCQCLNGACNTL 510
 DB 302 DCALRDCDNGACKEPBTGQCLCTAGMKNIKCDRPNCLNHFQGDCKAYVDCCHNNAACNP 361
 OY 511 DGTCTCAPGWRGECCELPQDGTGYLNGAERCDC--SHADGCHPTGHCRCPLGWSG 565
 DB 362 NSGCTCAAGWTGRCERKCDTGKFGHDCAKKCCQCCDFNNSLACATNGRCVCKODMG 418

RESULT 15
 ID 013149 PRELIMINARY; PRT; 2447 AA.
 AC 013149;
 DT 01-JUL-1997 (TREMBLrel. 04, Created)
 DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Notch 2 (Fragment).
 OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
 OC Acanthopterygii; Acanthopterygii; Perciformes; Tetraodontiformes;
 OC Tetraodontidae; Takifugu.
 OX NCBI_TaxID=31033;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Nakamura T., Townsdalet J.,
 RT "Structure of Fugu Notch gene."
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB004829; BAA20535.1; -.
 DR HSP; P00740; 1EDM.
 DR InterPro; IPR002110; ANK.
 DR InterPro; IPR000152; ASX_hydroxyl.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR000742; EGF_2.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR001438; EGF-II.
 DR InterPro; IPR000083; Fibinectn1.
 DR InterPro; IPR002049; Laminin_EGF.
 DR InterPro; IPR000800; Notch.
 DR Pfam; PF00023; ank; 6.
 DR Pfam; PF00008; EGF; 35.
 DR Pfam; PF00065; notch; 3.
 DR PRINTS; PR00010; EGFBLD.
 DR PRINTS; PR00011; EGF_LAMININ.
 DR PRINTS; PR00012; FNTYPEI.
 DR SMART; SM00248; ANK; 5.
 DR SMART; SM00179; EGF_CA; 20.
 DR SMART; SM00001; EGF_Like; 14.
 DR SMART; SM00004; NL; 3.
 DR PROSITE; PS50088; ANK_REPEAT; 8.
 DR PROSITE; PS50297; ANK_REPEAT_REGION; 2.
 DR PROSITE; PS00010; ASX_HYDROXYL; 22.
 DR PROSITE; PS00022; EGF_1; UNKNOWN_35.
 DR PROSITE; PS01187; EGF_2; 29.
 DR PROSITE; PS01186; EGF_CA; 20.
 KW ANK repeat; Calcium-binding; EGF-like domain; Glycoprotein;
 KW Hydroxylation; Repeat.
 FT NON_TER 1
 SO SEQUENCE 2447 AA; 262541 MW; 1A8E2A372A085D84 CRC64;

Query Match 15.2%; Score 1024.5; DB 13; Length 2447;
 Best local Similarity 26.3%; Pred. No. 8, 4e-74;
 Matches 320; Conservative 101; Mismatches 325; Indels 471; Gaps 70;

OY 6 NSCLSTFICLLCHWIGTASPLNLEDPNVCNHSWESYVQVESYV--HPFDQYVYTCSDI 63
 DB 99 DSCILSSPCAN--GGTCTSLT-----GGSTYCSCLPGTGRHCLNDTECATRS 145
 OY 64 L--NMFKCTRHRYSTFAYRHGERTMYRRKSQCCPFYEGSENC---VPRCADKCVHR 117
 DB 146 ICNNEGTCTINTRSY-----KCMCALGF--TKHCESSYIPCSPLNG 189
 OY 118 CLAPN-----TCCGEPGKNGTNCSSADGDHMRPHCTSRQCKNGALCNPTGACGACAG 172
 DB 190 TCNONSSTYSCHCLPGFNCTENNID--DCPHQCANRGTCTDGV---NTYNCQCPPE 244
 OY 173 FRGMRCEDECGTGYNDCH-QRCQCONGATCDHTGE--CRCPPTGAFCEP----- 223
 DB 245 WTGQCHTEDEV-----NECHLPQNTCONGTCNLTGSGYCYVYNWMSGLDCEENIDDD 298
 OY 224 -----LCPPGKHGPOC--EORC---PCONGGVV--HHVTG--ECSC 255

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 9, 2003, 04:45:58 ; Search time 8751 Seconds

(without alignments)
11383.717 Million cell updates/sec

Title: US-10-092-390-1

Perfect score: 3423

Sequence: 1 atggtattctcttctgaacac.....gcacgcagcagcagtgatga 3423

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenBank:*

1: gb_ba:*

2: gb_hg:*

3: gb_in:*

4: gb_om:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_sts:*

12: gb_sy:*

13: gb_un:*

14: gb_vi:*

15: em_ba:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_mu:*

20: em_om:*

21: em_or:*

22: em_ov:*

23: em_pat:*

24: em_ph:*

25: em_pl:*

26: em_ro:*

27: em_sts:*

28: em_un:*

29: em_vi:*

30: em_hg_hum:*

31: em_hg_inv:*

32: em_hg_other:*

33: em_hg_mus:*

34: em_hg_pln:*

35: em_hg_rtd:*

36: em_hg_mam:*

37: em_hg_vit:*

38: em_sy:*

39: em_hggo_hum:*

40: em_hggo_mus:*

41: em_hggo_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3421.4	100.0	7522	9	AB058676
2	1692.4	49.4	2267	9	BC020198
3	1200.6	35.1	5702	9	AB058677
4	742	21.7	3281	9	AL834326
5	649.4	19.0	4290	10	AF444274
6	648.6	18.9	4539	10	AF461685
7	646.2	18.9	4482	10	AF440279
8	548.2	16.0	660	10	RATGFRD
9	534.8	15.6	632	6	AK079681
10	456.8	13.3	4470	9	AK074121
11	400.8	11.7	5523	10	AB011532
12	354.8	10.4	4501	9	AB011539
13	250.8	7.3	146206	9	AC026800
14	250.8	7.3	152765	2	AC008566
15	250.8	7.3	175144	2	AC010415
16	250.8	7.3	217221	9	AC008682
17	243	7.1	48202	2	AC012911
18	243	7.1	172175	3	AC010038
19	243	7.1	177583	3	AC105264
20	243	7.1	257867	3	AC005557
21	243	7.1	303191	3	AE003472
22	241.8	7.1	192282	9	AC010424
23	239.8	7.0	2054	3	AK098809
24	216	6.3	3764	3	AF332568
25	197.2	5.8	163437	2	AC102173
26	192.4	5.6	153960	2	AC095558
27	192.4	5.6	178298	2	AC094759
28	180	5.3	62877	2	AC104414
29	166	4.8	178298	2	AC094759
30	153	4.5	75342	3	CEY479C
31	149	4.4	73791	2	AC091031
32	149	4.4	169638	9	AC009433
33	149	4.4	185854	2	AC027736
34	147.4	4.3	178862	9	AC011847
35	145.2	4.2	153960	2	AC095558
36	140.2	4.1	177355	2	AC106503
37	139.4	4.1	230668	2	AC121921
38	139.4	4.1	233204	10	AC092498
39	130.6	3.8	405	11	G51780
40	119.6	3.5	87428	3	CEY64G10A
41	119.6	3.5	297900	2	AC006704
42	118.8	3.5	73791	2	AC091031
43	118.2	3.5	3210	3	AF145617
44	117.8	3.4	152765	2	AC008566
45	117.8	3.4	155630	2	AC015721

ALIGNMENTS

RESULT 1

AB058676

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Homo sapiens adult. hippocampus cDNA to mRNA, clone:pf01012.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (sites)

Nagase,T., Nakayama,M., Nakajima,D., Kikuno,R. and Ohara,O.

Prediction of the coding sequences of unidentified human genes. xx.

The complete sequences of 100 new cDNA clones from brain which code

Prod. No. is the number of results predicted by chance to have a

for large proteins in vitro
 JOURNAL DNA Res. 8 (2), 85-95 (2001)
 MEDLINE 21245130
 2 (bases 1 to 7522)
 AUTHORS Nakayama, M., Nagase, T., Nakajima, D., Kikuno, R. and Ohara, O.
 TITLE Direct Submission
 JOURNAL Submitted (27-MAR-2001) Manabu Nakayama, Kazusa DNA Research
 Institute, Department of Human Gene Research, 1532-3, Yama,
 Kisarazu, Chiba 292-0812, Japan (E-mail: manabukazusa.or.jp,
 URL: http://www.kazusa.or.jp/huge, Tel: 81-438-52-3915,
 Fax: 81-438-52-3914)

FEATURES

Source
 Location/Qualifiers
 1..7522
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="pf01012"
 /tissue_type="hippocampus"
 /dev_stage="adult"
 /note="vector:pluescript II SK plus"
 1..7522
 /gene="MEGF10"
 204..3626
 /gene="MEGF10"
 /note="KIAA1780 protein
 gene encoding protein with multiple EGF-like-domains"
 /codon_start=1
 /product="MEGF10 protein (KIAA1780)"
 /protein_id="BAB47409.1"
 /db_xref="GI:1401777"
 /translation="MVISNLSLFTICLLIHWIGTASPLNEDPNCSSHMSYSVT
 DESYPHPDQIYYTCTDIPLNFKCTRIHRSYRTYRIGETWTRKSKQCKYSEB
 EMCVPHCKADCKYHRCIAPNTQCEPFGWGTSSACAGDHDHGPCTSRQCKNALC
 NPITGACHCAAFGRMRCEDRCEDGTGNDCHQRCOCONGATCDVTEBCRPDYTG
 AFCEIDCPKRGHPQCEQRCPCONGVCHHTVDECSGMMGYVCGAPCPGRFGKN
 CSQECQNHNGTCDATAGCCHSPGTGRCODEPVGTYGLCAETGEMSCYAGKCYH
 VSGACIENGFAGRGCEPARIPEGLYGIKCKRRCCHLEHTHSHPMGEACCKRGSY
 GLYKCHTSRPFYGRACQOICSCONGACDSYTGCTCAPEGKIGDCSTPCPLGTGTGI
 NCSSRGCKNDVCSFVDSCTCKAGMHWVDSILKPSGTWFGNLTQCLNGACN
 TLDGTCTAPGRWGERCELPQDGTGYNCAERCCSHADGCHPTTGRCILPMSGV
 HCDVYACGRMWPNCSLPCYKNGASCSDDGICBAGFRGTTCORLCSPEFYHRC
 SPTQPCVHSSGPHHITGICDLPLGFTGALCNVCPGSRFGKNCAGICTCTNNKCN
 PIDRSCOCYPGIGSDGSCPCPAMHGNICHTGCHNGAFCSAYDGEKCTPGMTGL
 VCTORCPLEFGYKCDALICCONGADCHISGOCCTCRGPFMRHCEQCPSTYKGC
 ROICDCLNNSTCDHTITGTCYCSPEKKGACRDAGIYIGNLSLSTRTALPADSTQI
 GAIAGIILVLEVLALFTIYRHQKQKESMSAVYTPMAYRVNADYITISGLPH
 SNGNANSHYFTNPSTHTLTCATSPHNNRDMVTYSKNNQLPVLKNVNGRGP
 VEDGTGLPADMKHGGYLNELGAFGLDRSYMKSJDLKNSYNSNCSLSSSENPY
 ATIKDPVLIPEKSECGYEMKSPARDSPEIINNSTSANNVYEVEPTVAVGVF
 SNMRLSDPPVDLPKNSHPCHYDLIPVADSSSPKODSGSSSSSSSSSP"

BASE COUNT 2130 a 1554 c 1697 g 2141 t
 ORIGIN

Query Match 100.0%; Score 3421.4; DB 9; Length 7522;
 Best Local Similarity 100.0%; Pred. No. 0;

Matches 3422; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGGTATTCTTTTGAACATGCGCTGAGCTTTATTTGTTATGTTATGCCACTGGATT 60
 DB 204 ATGGTATTCTTTGAAACATGCGCTGAGCTTTATTTGTTATGTTATGCCACTGGATT 263
 QY 61 GGGACAGCATCACTGATCTTGAAGACCCCTAATGTGTGACCACTGGGAAACCTAC 120
 DB 264 GGGACAGCATCACTGATCTTGAAGACCCCTAATGTGTGACCACTGGGAAACCTAC 323
 QY 121 TCAGTACTGTGCAGAGATCATACCCATCCCTTGTGATCAATTTACTACAGACTGC 180
 DB 324 TAGAGTACTGTGCAGAGATCATACCCATCCCTTGTGATCAATTTACTACAGACTGC 383
 QY 181 ACTGACATTTAAACTGTTTAAATGACAGCGGACAGAGTCACTATCGGACAGCTAT 240
 DB 384 ACTGACATTTAAACTGTTTAAATGACAGCGGACAGAGTCACTATCGGACAGCTAT 443
 QY 241 CGACATGGGGAGAAAGACTATGATAGGCGCAAGTCTCATGTTGTCTCGATTTTATGAA 300

DB 444 CGACATGGGGAGAAAGACTATGATAGGCGCAAGTCTCATGTTGTCTCGATTTTATGAA 503
 QY 301 AGCGGGAAATGTGTGTCCCCCACTGTGCTGATTAATGTGTCCATGTGTGGCTATGCT 360
 DB 504 AGCGGGAAATGTGTGTCCCCCACTGTGTGATTAATGTGTCCATGTGTGGCTATGCT 563
 QY 361 CCAAACACCTGTACGTGTAGAGCTGTGGTGGGAGGACCAACTGTCTCAGTCCGCTGAT 420
 DB 564 CCAAACACCTGTACGTGTAGAGCTGTGGTGGGAGGACCAACTGTCTCAGTCCGCTGAT 623
 QY 421 GGTGATCACTGGGGTCCCACTGTACACAGCCGGTGCAGTGCAGAAATGGGCTGTGTC 480
 DB 624 GGTGATCACTGGGGTCCCACTGTACACAGCCGGTGCAGTGCAGAAATGGGCTGTGTC 683
 QY 481 AACCCCATCACGGGGCTTGCACATGTGTGGGGGCTTCCGGGGCTGGGGCTGGCGAGAC 540
 DB 684 AACCCCATCACGGGGCTTGCACATGTGTGGGGGCTTCCGGGGCTGGGGCTGGCGAGAC 743
 QY 541 CGCTGTAGCAGGAGGACCTATGTGTAACGACTGTATCAGATGATGCCAGTCCAGATGGA 600
 DB 744 CGCTGTAGCAGGAGGACCTATGTGTAACGACTGTATCAGATGATGCCAGTCCAGATGGA 803
 QY 601 GCCACCTGGACACGCTCACGGGGGAATGCCGTGCCACACAGATACACCGGAGCTTTC 660
 DB 804 GCCACCTGGACACGCTCACGGGGGAATGCCGTGCCACACAGATACACCGGAGCTTTC 863
 QY 661 TGTGAGATCTTGTCTCTCTGTTAAACATGTGTCACAGTGTAGAGAGATGCCCTTGT 720
 DB 864 TGTGAGATCTTGTCTCTCTGTTAAACATGTGTCACAGTGTAGAGAGATGCCCTTGT 923
 QY 721 CAAATGAGAGAGTGTGTCATACGTCACGTGAGAAATGCTTCCCTTCTGGCTGATG 780
 DB 924 CAAATGAGAGAGTGTGTCATACGTCACGTGAGAAATGCTTCCCTTCTGGCTGATG 983
 QY 781 GGCACAGTGTGTGTAGCTTGGCCCCGAGGGTGGCTTGGAAAGACTTCCCAAGAA 840
 DB 984 GGCACAGTGTGTGTAGCTTGGCCCCGAGGGTGGCTTGGAAAGACTTCCCAAGAA 1043
 QY 841 TGCACATGTCATTAATGAGAGGAGTGTGATGCTGACACAGGCAATGTATTCAGTCA 900
 DB 1044 TGCACATGTCATTAATGAGAGGAGTGTGATGCTGACACAGGCAATGTATTCAGTCA 1103
 QY 901 GGTATACACAGGGAGACGGTGTCCAGATGATGTCTGTTGGACCTATGAGCGTTCTGT 960
 DB 1104 GGTATACACAGGGAGACGGTGTCCAGATGATGTCTGTTGGACCTATGAGCGTTCTGT 1163
 QY 961 GGTGAGACCTGTGCAGTGTGCAACGAGGAGAGTGTACACGTGAGGGGCGATGCCTC 1020
 DB 1164 GGTGAGACCTGTGCAGTGTGCAACGAGGAGAGTGTACACGTGAGGGGCGATGCCTC 1223
 QY 1021 TGTGAAGCAGGGCTTGTGCTGGCGAGCGCTCGAAGACGCGCTGTCTTGAAGGGCTCTAC 1080
 DB 1224 TGTGAAGCAGGGCTTGTGCTGGCGAGCGCTCGAAGACGCGCTGTCTTGAAGGGCTCTAC 1283
 QY 1081 GGCATCAATGTGACAAACGGTGTCTCTCCACCTTGGAAAAACATCATATAGCTACCCC 1140
 DB 1284 GGCATCAATGTGACAAACGGTGTCTCTCCACCTTGGAAAAACATCATATAGCTACCCC 1343
 QY 1141 ATGTCTGAGAGTGTGCTGTGCAAGCGGGCTGTGCAAGTGTGCAAGTGTGCAAGTGT 1200
 DB 1344 ATGTCTGAGAGTGTGCTGTGCAAGCGGGCTGTGCAAGTGTGCAAGTGTGCAAGTGT 1403
 QY 1201 TCTCCTGATTTACAGGGGAAGCTTGCACAGATGTGCAAGTGTGCAAGTGTGCAAGTGT 1260
 DB 1404 TCTCCTGATTTACAGGGGAAGCTTGCACAGATGTGCAAGTGTGCAAGTGTGCAAGTGT 1463
 QY 1261 TGTGACAGTGTGACAGAAAGTGTACCTGTGCCCGACAGATTAAGAAATGACCTGCT 1320
 DB 1464 TGTGACAGTGTGACAGAAAGTGTACCTGTGCCCGACAGATTAAGAAATGACCTGCT 1523
 QY 1321 ACCCATGCGCTTGGGAAGCTATGGGATTAACCTGTCTCTGCTGTGGCTGTAAAT 1380

Db	1524	ACCCCATGCCCTCTGGGAAACGTAATGGGATAAAGTCTTCCTCTGGCTGTGGCTGTATAAAT	1583
Qy	1381	GATGAGTCTGTCTCTCTGTGTGGAGGGGTCTGTACTTTCGAAGGACAGCTGGGACGGGGTG	1440
Db	1584	GATCAGCTGTCTCTCTGTGTGGAGGGGTCTGTACTTTCGAAGGACAGCTGGGACGGGGTG	1643
Qy	1441	GACTGTCCATCAGATGTCCACAGTGGGACATGGGGGCTTTGGCGTGAATTAACATGCCAG	1500
Db	1644	GACTGTCCATCAGATGTCCAGTGGGACATGGGGGCTTTGGCGTGAATTAACATGCCAG	1703
Qy	1501	TGCTCAACGGGGGAGGCTGCACACCCCTGGAGGGGACCTGCACGCTGTCCACTGGATGG	1560
Db	1704	TGCTCAACGGGGGAGGCTGCACACCCCTGGAGGGGACCTGCACGCTGTCCACTGGATGG	1763
Qy	1561	CGCGGGGAGAAATCGAAATCTCCCTGGCAGAGATGGCAGTGAAGCTGTGTGG	1620
Db	1764	CGCGGGGAGAAATCGAAATCTCCCTGGCAGAGATGGCAGTGAAGCTGTGTGG	1823
Qy	1621	CGCTGCACATGCACGACGAGATGGGTGCACCTTACACGGGGCATTTGCCGTCTC	1680
Db	1824	CGCTGCACATGCACGACGAGATGGGTGCACCTTACACGGGGCATTTGCCGTCTC	1883
Qy	1681	CCGGGATGTGTAGTGTTCACATGTGACAGCTGTGTCTGAGGGAGCGTGGGGCCCCAAC	1740
Db	1884	CCGGGATGTGTAGTGTTCACATGTGACAGCTGTGTCTGAGGGAGCGTGGGGCCCCAAC	1943
Qy	1741	TGCTCCGCGCCCTCTCTGTAAAAATGGGGCTTCATGGCTCCCTCATATATGGCATCTGC	1800
Db	1944	TGCTCCGCGCCCTCTCTGTAAAAATGGGGCTTCATGGCTCCCTCATATATGGCATCTGC	2003
Qy	1801	GAGTGTGCACACGAGCTTCGAGGACACCTTGTTCAGAGATTCGTCCTCCCTGGTTTTAT	1860
Db	2004	GAGTGTGCACACGAGCTTCGAGGACACCTTGTTCAGAGATTCGTCCTCCCTGGTTTTAT	2063
Qy	1861	GGGGATGTGTGCACGACGAGATGCCAGTGGGTTCACAGGAGGGGGCCCTGCCACAC	1920
Db	2064	GGGGATGTGTGCACGACGAGATGCCAGTGGGTTCACAGGAGGGGGCCCTGCCACAC	2123
Qy	1921	ATCACGGGCTGTGTGACTCTTGCCTGGGCTTCACAGGCGCCCTCTGCATATGAGTGT	1980
Db	2124	ATCACGGGCTGTGTGACTCTTGCCTGGGCTTCACAGGCGCCCTCTGCATATGAGTGT	2183
Qy	1981	CCGAGTGGCAGATTTGGGAAAAACGTGTGCAGAAATTTGTACTGTGCACCAACGAGAAC	2040
Db	2184	CCGAGTGGCAGATTTGGGAAAAACGTGTGCAGAAATTTGTACTGTGCACCAACGAGAAC	2243
Qy	2041	TGTAAACCCATTTGACATATCTTGTACGTGTACCCGGTGTGATTTGGAGTGCCTCT	2100
Db	2244	TGTAAACCCATTTGACATATCTTGTACGTGTACCCGGTGTGATTTGGAGTGCCTCT	2303
Qy	2101	CAACCATGTCCACCTGCCCACTGGGGCCCAACTGCATCCACAGTGCACATGCCATAT	2160
Db	2304	CAACCATGTCCACCTGCCCACTGGGGCCCAACTGCATCCACAGTGCACATGCCATAT	2363
Qy	2161	GGAGCTTTGTGCACGCTGACGATGGGGGATGTAAATGACCTCTGGCTGGAGCAGGGCTC	2220
Db	2364	GGAGCTTTGTGCACGCTGACGATGGGGGATGTAAATGACCTCTGGCTGGAGCAGGGCTC	2423
Qy	2221	TACTGCACCTAGAGATGTCTCTAGGGTTTTATGAAAAAGATTTGTCACTGTATGCCAA	2280
Db	2424	TACTGCACCTAGAGATGTCTCTAGGGTTTTATGAAAAAGATTTGTCACTGTATGCCAA	2483
Qy	2281	TGTCAAAAAGGAGTGTGACCTGCAGACCATTTCTGGGCAAGTACTTGGCCGACTGGATT	2340
Db	2484	TGTCAAAAAGGAGTGTGACCTGCAGACCATTTCTGGGCAAGTACTTGGCCGACTGGATT	2543
Qy	2341	ATGGGAGGCACTGTGACGAGAAAGTGCCTTCAGAAACATATGGCTATGGCTGTGCCAG	2400
Db	2544	ATGGGAGGCACTGTGACGAGAAAGTGCCTTCAGAAACATATGGCTATGGCTGTGCCAG	2603
Qy	2401	ATATGTATTTGTCTGAACCACTCCACTGTGCACCATCACTGGGACCTGTATCTGCAGC	2460
Db	2604	ATATGTATTTGTCTGAACCACTCCACTGTGCACCATCACTGGGACCTGTATCTGCAGC	2663
Qy	2461	CCCGATGGAAGGAGCGGAGATGTGATCAAGCTGTGTATATCATATGTTGAAATTTGAAC	2520
Db	2664	CCCGATGGAAGGAGCGGAGATGTGATCAAGCTGTGTATATCATATGTTGAAATTTGAAC	2723
Qy	2521	AGCTTAAGCCGAACACTACTGTGCTCCCGCTGGATTCCTACCAAGATCGGGCCATTTGCA	2580
Db	2724	AGCTTAAGCCGAACACTACTGTGCTCCCGCTGGATTCCTACCAAGATCGGGCCATTTGCA	2783
Qy	2581	GGCATCATCTTCTGTCTAGTGTCTCTCTCCACTGGCATTTGTATTTATAGA	2640
Db	2784	GGCATCATCTTCTGTCTAGTGTCTCTCTCCACTGGCATTTGTATTTATAGA	2843
Qy	2641	CACAAAGAAAGGAAAGAAATCAACGATGCCAGAGTTTACTTACACCCCTCTATGAG	2700
Db	2844	CACAAAGAAAGGAAAGAAATCAACGATGCCAGAGTTTACTTACACCCCTCTATGAG	2903
Qy	2701	GTCTCATATCAGATTTATACATTTGAGAAACCTTCTCCACAGCAATGTGGAAACGT	2760
Db	2904	GTCTCATATCAGATTTATACATTTGAGAAACCTTCTCCACAGCAATGTGGAAACGT	2963
Qy	2761	AATAGCCACTACTTACCAATTCACAGTTTACACACGCTACACCGAGTGTGCCATCCCT	2820
Db	2964	AATAGCCACTACTTACCAATTCACAGTTTACACACGCTACACCGAGTGTGCCATCCCT	3023
Qy	2821	CAGCTCAACAAAGGGAGCAGATGACGTGCAGCAAGTCAAAAAACAATCACTGTTTTG	2880
Db	3024	CAGCTCAACAAAGGGAGCAGATGACGTGCAGCAAGTCAAAAAACAATCACTGTTTTG	3083
Qy	2881	AATCTTAAAAATGTGAACCTGTGGGAAGAGGCCCTGTGGGGACTGTCACTGGGCAATTG	2940
Db	3084	AATCTTAAAAATGTGAACCTGTGGGAAGAGGCCCTGTGGGGACTGTCACTGGGCAATTG	3143
Qy	2941	CCGGCTACTGTGAACATGGGGGCTACCTCAAGACACTGGGTCTTTGGATTGACGA	3000
Db	3144	CCGGCTACTGTGAACATGGGGGCTACCTCAAGACACTGGGTCTTTGGATTGACGA	3203
Qy	3001	AGCTATATGGGAAAAATCCTTAAAAAGACCTGGGAAAAATTCGTGAATATATTTCAAGTAA	3060
Db	3204	AGCTATATGGGAAAAATCCTTAAAAAGACCTGGGAAAAATTCGTGAATATATTTCAAGTAA	3263
Qy	3061	TGCTCCCTAAGCAGTTCTGGAACCCATATGCCACTATTTAAGACCCACTGTACTTATC	3120
Db	3264	TGCTCCCTAAGCAGTTCTGGAACCCATATGCCACTATTTAAGACCCACTGTACTTATC	3323
Qy	3121	CCGAAAGCTCAGAGTGTGTATGTGAGATGAATTCGCCGACGAGAGATTTCCCA	3180
Db	3324	CCGAAAGCTCAGAGTGTGTATGTGAGATGAATTCGCCGACGAGAGATTTCCCA	3383
Qy	3181	TATGCAGAGATCAATTAATCACTCACTCAGCCACAGAGATGTCTATGAAGTTGAACCTTACA	3240
Db	3384	TATGCAGAGATCAATTAATCACTCACTCAGCCACAGAGATGTCTATGAAGTTGAACCTTACA	3443
Qy	3241	GTGAGTGTGTGCCAAGGAGTATTCAGCAATTAATGGGGGCTCTGCCAGATCCATATGAC	3300
Db	3444	GTGAGTGTGTGCCAAGGAGTATTCAGCAATTAATGGGGGCTCTGCCAGATCCATATGAC	3503
Qy	3301	CTCCCAAGAAAGTCACATCCCTGTCTATATGACCTCTGCGACCTCGAGACAGTTCA	3360
Db	3504	CTCCCAAGAAAGTCACATCCCTGTCTATATGACCTCTGCGACCTCGAGACAGTTCA	3563
Qy	3361	TTCCTCCCTAAGCAAGGAGCACTGGAAGTGAAGAGACGAACAGCAAGCAGCAGGTGA	3420
Db	3564	TTCCTCCCTAAGCAAGGAGCACTGGAAGTGAAGAGACGAACAGCAGCAGCAGGTGA	3623
Qy	3421	TGA 3423	
Db	3624	TGA 3626	
RESULT 2			
BC020198			
LOCUS	BC020198	2267 bp	mRNA linear PRI 22-JAN-2002

DEFINITION Homo sapiens, similar to MGC10 protein, clone MGC:31776
IMAGE:4904255, mRNA, complete cds.
ACCESSION BC020198
VERSION BC020198.1 GI:18044365
KEYWORDS MGC.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 2267)
REFERENCE Strausberg, R.
AUTHORS Direct Submission
TITLE Submitted (19-DEC-2001) National Institutes of Health, Mammalian
JOURNAL Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
COMMENT Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Genome Sequence Centre,
BC Cancer Agency, Vancouver, BC, Canada
info@cgsc.bc.ca
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
Leticia Hailo, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
Ness, Pawan Pandoh, Anna-Lisa Prabhu, Parvaneh Saeedi, Jacqueline
Schell, Duane Smilans, Michael Smith, Lorraine Speede, Jeff Stott,
Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
George Yang, Scott Zuyderduyn, Marco Marra.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: <http://Image.lnl.gov>
Series: IRAL Plate: 40 Row: 9 Column: 4
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 14192942.
FEATURES
source
location/Qualifiers
1..2267
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MGC:31776 IMAGE:4904255"
/tissue_type="Muscle, Fibroblastsarcoma"
/clone_id="NH_MGC_17"
/lab_host="DH10B-R"
/note="Vector: pOTB7"
280..1983
/codon_start=1
/product="Similar to MGC10 protein"
/protein_id="AAH20198.1"
/db_xref="gi:18044366"
/translation="MVISLNSCLSFICLLCHMIGTASPLNLEDPNVCSSHMSYSVT
OESYHPFDIYTSCTDILNMFKTRHVSRTYRHYRGEKMYRRKSOCCPFESG
EMCVPHCADKCYHRCIAIPNTCCCEPFGGNTSSACGSDHHPCTSRCCCKNALC
NPITGACAGAGFRGWRCEDEBOGTYNNDHQRCQCNNGATCDVTEBCRPPEYTG
AFEDLCPGKAGHPQCEQRCPOQNGVYCHVHYGSCSPGNGWTVCGPCEPGRGKN
CSOECCHNGGTCDNAEQCHCSPELTGERODECPVTVGLCAETQCQVAGKCYN
VSGALCEAGFAGEERCEARLCEPGLYGIKDCRCHLENTSHCHPMGSECKRFGMS
GLYENETSPPFYGEACQICSCONGACDDSTGCTCAAPGKGTGDCSTPCPLGTGCI
NCSRCCKKNANVCSPVDSCTCRKAGHGVDSINCPGRTMFGCNLTGCCINGACN
TLDTGTCAPGMRGEKELPCODGTGGLNCAERCDKSHADGCHPTTGHCRLPGWSGL
F"

BASE COUNT 543 a 548 c 631 g 545 t
ORIGIN
Query Match 49.4%; Score 1692.4; DB 9; Length 2267;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1693; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 AAGGATATTCTTGAACGATCGGACCTTATTTGTTATGTTATGACCACTGATTT 60
|||||

Db 280 AAGGATATTCTTGAACGATCGGACCTTATTTGTTATGTTATGACCACTGATTT 339
QY 61 GGGACAGCATCACTGATGATCTTGAAGACCCATATGTGTAGCCACTGGGAAAGCTAC 120
|||||
Db 340 GGGACAGCATCACTGATGATCTTGAAGACCCATATGTGTAGCCACTGGGAAAGCTAC 399
QY 121 TCAGTACGTGTGCAGAGATCATACCCACATCCCTTTGATCAATTTACTACAGAGCTGC 180
Db 400 TCAGTACGTGTGCAGAGATCATACCCACATCCCTTTGATCAATTTACTACAGAGCTGC 459
QY 181 ACTGACATTTAACTGCTTTAAATATCAGCGGCAAGAGTACGATATGGGACAGCTAT 240
Db 460 ACTGACATTTAACTGCTTTAAATATCAGCGGCAAGAGTACGATATGGGACAGCTAT 519
QY 241 CGACATGGGAGAGAGACTATGATATAGCGCAAGTCACTGTGCTGCTGATTTATGAA 300
Db 520 CGACATGGGAGAGAGACTATGATATAGCGCAAGTCACTGTGCTGCTGATTTATGAA 579
QY 301 AGCGGGGAAATGTGTCTCCCACTGCTGATTAATGTGTCCATGCTGTGATTTGCT 360
Db 580 AGCGGGGAAATGTGTCTCCCACTGCTGATTAATGTGTCCATGCTGTGATTTGCT 639
QY 361 CCAAAACCTGTCAGTGTAGCTGCTGCTGGGGAGGAGACCACTGCTCAAGTCCGAT 420
Db 640 CCAAAACCTGTCAGTGTAGCTGCTGCTGGGGAGGAGACCACTGCTCAAGTCCGAT 699
QY 421 GGTGATCACTGGGGTCCCACTGACACAGCCGCTGCGACAAAATGGGGCTGTGTC 480
Db 700 GGTGATCACTGGGGTCCCACTGACACAGCCGCTGCGACAAAATGGGGCTGTGTC 759
QY 481 AACCCATACCGGGGCTTCCACTGCTGCTGGGGCTTCCGGGGCTGGGGCTGGAGAC 540
Db 760 AACCCATACCGGGGCTTCCACTGCTGCTGGGGCTTCCGGGGCTGGGGCTGGAGAC 819
QY 541 CGCTGTGACAGGGGACCTATGATGATACAGTGTGATCAAGATGCCAGATGGA 600
Db 820 CGCTGTGACAGGGGACCTATGATGATACAGTGTGATCAAGATGCCAGATGGA 879
QY 601 GCCACCTGGACACGTCACAGGGGGAATGCCGTGCCACACAGATACACAGGAGCTTC 660
Db 880 GCCACCTGGACACGTCACAGGGGGAATGCCGTGCCACACAGATACACAGGAGCTTC 939
QY 661 TGTGAGATCTTGTCTCTCTGTTAAACATGTGTCCACATGTGTGAGAGATGCCCTTGT 720
Db 940 TGTGAGATCTTGTCTCTCTGTTAAACATGTGTCCACATGTGTGAGAGATGCCCTTGT 999
QY 721 CAAAATGAGAGATGTGTATCATCATCATGTGAGATGCTTGGCTTGGCTGGATG 780
Db 1000 CAAAATGAGAGATGTGTATCATCATCATGTGAGATGCTTGGCTTGGCTGGATG 1059
QY 781 GGCACAGTGTGTGTGACGCTTGCCTGAGAGAGGTGCTTGGAAAGATGTTCCCAAGAA 840
Db 1060 GGCACAGTGTGTGTGACGCTTGCCTGAGAGAGGTGCTTGGAAAGATGTTCCCAAGAA 1119
QY 841 TGGCATGTGCTTAATGAGAGAGATGTGATGCTGCGCACAGGCCAATGTATTCATGCCA 900
Db 1120 TGGCATGTGCTTAATGAGAGAGATGTGATGCTGCGCACAGGCCAATGTATTCATGCCA 1179
QY 901 GGATACACAGGGGAAAGGAGTGCAGATGATGATGCTGTTGGGACATATGGGCTTCTCT 960
Db 1180 GGATACACAGGGGAAAGGAGTGCAGATGATGATGCTGTTGGGACATATGGGCTTCTCT 1239
QY 961 GGTGAGACCTGTCAGTGTGTCAACGAGAGGAAGTGTTCACAGCTGAGCGGCGATGCC 1020
Db 1240 GGTGAGACCTGTCAGTGTGTCAACGAGAGGAAGTGTTCACAGCTGAGCGGCGATGCC 1299
QY 1021 TGTGAAGCAGGCTTGTGCTGCGAGCGCTGCGAAGCAGCCTGTGTCTAAGGGGCTCTAC 1080
Db 1300 TGTGAAGCAGGCTTGTGCTGCGAGCGCTGCGAAGCAGCCTGTGTCTAAGGGGCTCTAC 1359
QY 1081 GGCATCAATGTGACAAAGGAGTGTGCTGCGCACTTGGAAACATCATATGCTGTACCCC 1140
Db 1360 GGCATCAATGTGACAAAGGAGTGTGCTGCGCACTTGGAAACATCATATGCTGTACCCC 1419

```

QY 1141 ATGCTGAGAGTGTGCTGCAAGCCGGGCTGGTCCAGGACTCTACTGTATGACATGT 1200
DB 1420 ATGTCTGGAGAGTGTGCTGCAAGCCGGGCTGGTCCAGGACTCTACTGTATGACATGT 1479
QY 1201 TCTCTGAGATTCACGGGGAAGCTTGCACAGATCTGACAGTCCCAAAATGGGGGAC 1260
DB 1480 TCTCTGAGATTCACGGGGAAGCTTGCACAGATCTGACAGTCCCAAAATGGGGGAC 1339
QY 1261 TGTGACAGTGTACGTGAAAGTGCACCTGTGCCCCAGAGATTCAAGGAATGACTGCT 1320
DB 1540 TGTGACAGTGTACGTGAAAGTGCACCTGTGCCCCAGAGATTCAAGGAATGACTGCT 1599
QY 1321 ACCCATGCCCCCTGGGAACTTATGGGATTAATGTTCTCTGCTGCTGTAAATAAT 1380
DB 1600 ACCCATGCCCCCTGGGAACTTATGGGATTAATGTTCTCTGCTGCTGTAAATAAT 1659
QY 1381 GATGCAAGTCTGCTCTCTGTGAGAGGGTCTTGTACTTGAAGCAGAGCTGGACGGGGT 1440
DB 1660 GATGCAAGTCTGCTCTCTGTGAGAGGGTCTTGTACTTGAAGCAGAGCTGGACGGGGT 1719
QY 1441 GACTGCTCCATCAGATGTCCCAAGTGGCAGCATGGGGCTTTGGCTGTAACTTAACATGCCAG 1500
DB 1720 GACTGCTCCATCAGATGTCCCAAGTGGCAGCATGGGGCTTTGGCTGTAACTTAACATGCCAG 1779
QY 1501 TGCCTCAAGGGGGAGGCTGCAACACCCCTGGAGGGAGCTGCAAGTGTGACCTGGATGG 1560
DB 1780 TGCCTCAAGGGGGAGGCTGCAACACCCCTGGAGGGAGCTGCAAGTGTGACCTGGATGG 1839
QY 1561 CGCGGGGAGAAATGCGAACTTCCCTGCGCAGAGTGGCAGCATGGGGCTTGAAGTGTGAG 1620
DB 1840 CGCGGGGAGAAATGCGAACTTCCCTGCGCAGAGTGGCAGCATGGGGCTTGAAGTGTGAG 1899
QY 1621 CGCTGGCAGTGCAGCCAGCAGATGGTGTGCTACCTTACACGGGCTATTCGGCTGCTC 1680
DB 1900 CGCTGGCAGTGCAGCCAGCAGATGGTGTGCTACCTTACACGGGCTATTCGGCTGCTC 1959
QY 1681 CCGGATGTCTCAGG 1694
DB 1960 CCGGATGTCTCAGG 1973

LOCUS AB058677 5702 bp mRNA linear PRI 10-MAY-2001
DEFINITION Homo sapiens mRNA for MEGF11 protein (K1A1781), complete cds.
ACCESSION AB058677
VERSION AB058677.1 GI:14017778
KEYWORDS
SOURCE Homo sapiens
ORGANISM Homo sapiens fetal brain cDNA to mRNA, clone:fg06971.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
1 (sites)
Nagase, T., Nakayama, M., Nakajima, D., Kikuno, R. and Ohara, O.
Prediction of the coding sequences of unidentified human genes. XX.
The complete sequences of 100 new cDNA clones from brain which code
for large proteins in vitro
DNA Res. 8 (2), 85-95 (2001)
JOURNAL MEDLINE
2 (bases 1 to 5702)
21245130
REFERENCE
Nakayama, M., Nagase, T., Nakajima, D., Kikuno, R. and Ohara, O.
Direct Submission
JOURNAL Submitted (27-MAR-2001) Manabu Nakayama, Kazusa DNA Research
Institute, Department of Human Gene Research, 1532-3, Yama,
Kisarazu, Chiba 292-0812, Japan (E-mail:manabu@kazusa.or.jp,
URL:http://www.kazusa.or.jp/huge, Tel:81-438-52-3915,
Fax:81-438-52-3914)
FEATURES
Source
Location/Qualifiers
1..5702
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="fg06971"

```

```

/tissue-type="brain"
/dev_stage="fetal"
/notes="vector:pluascript II SK plus"
1..5702
/gene="MEG11"
160..3069
/gene="MEG11"
/notes="K1A1781 protein
Start codon is not confirmed. fg06971 cDNA clone for
K1A1781 has a 119bp insertion after the position 2642 of
the sequence of K1A1781
gene encoding protein with multiple EGF-like domains"
/codon_start=1
/product="MEG11 protein (K1A1781)"
/protein_id="BAB47410.1"
/db_xref="GI:14017779"
/translation="MHTPSIRSIHTDAQTSSTSSNAPGATLCTERCVRGCVSPDTC
CEPMNGGPDSSGSDSDHMGPHSCNRCOMCALNPITGACVCAAGFPMGCEELCA
PGTHGKGLPCQCRHAGSCDPRAGRLCAPYTGVEELCPGSHGHELCRPGCQ
NGCTGHTHTGECACPRGWTGAYCAACPCPGCTGGONGSDCPCHHGGGCDHYGQCHT
ATYMDROCEBPFPSFGQCSQRCHNHGGCSPFTGCECEPEYKRPCEDERLCPE
GLHGECDLPCEADADNTISCHPVTAAGTCQPGMGSHHNECSPVGYTDGQLPCTC
ONGADCHSTTGCTCAGPFMGVCAVSCAAGTYGPNCSISGNNGTCSVPDAGSCPT
KEGMGLDCTLPSPGWTGLNENSCCANAGATSIDSCSTPGMLDTCLEPCD
GTFGLNCSHCCDASHADGCDPVTGHCCCAWGTGIRJCDSTCPGRMGPNCSYSCEN
GGSCSPEDSCBACBGRERGLRQICRDPPEFYHGGAGCPCLVHSSRPCHHSIGICE
LPGFSALCNQYACAGYREGDDAQLCSANNCTSPIDSCCCEPFMGIKDCSONCP
LPGFACPFACSHONGASCADGACHCTPGWTGLCTDRCRAFFGKDCGVCOON
GASCHDISKCTRTGFTGQHEORCAPPTFGYGCQQLCEKNNSTCHVTCTVCSF
GFGICRDLAALAMELEINPYTKISPLAGEHSVAGVIMLFFIYVLGLFMHR
RQKEKGRDLAPRVSTYTPAMRMTSTYSLGACGMDRQNTYIMDKGFDYKESYS
STGCSLNSSENENYATIKDPIILCKLPSESYVEMKSPVMGSPYTDVPSLSTNNIY
EVEPTSYVQBECCGNHNSYIQNAYDLPFRSHLPGRHDLPLVQSPANPESQDKOS"
BASE COUNT 1495 a 1500 c 1382 g 1325 t
ORIGIN
Query Match 35.1%; Score 1200.6; DB 9; Length 5702;
Best Local Similarity 69.2%; Pred. No. 0;
Matches 1670; Conservative 0; Mismatches 739; Indels 6; Gaps 2;
QY 318 CCCCACGTGTGTATTAATGTGTCCATGCTGCTATGTCCTCAACACCTGTCACTG 377
DB 234 CGCCCTGTGTATCAGGAGAGTGTGTGACAGCCGCTGGCTTCCCGGACACTGTGCACTG 293
QY 378 TGAGCCTGTGTGGGAGAGACCAATGCTCTCAATGCTGCTGAGATGTACTGGGGTCC 437
DB 294 CGAGCCTGTGTGGGAGAGGCCCGGACGCTGCTCAACGCGCTGTGGAACGAGCACCTGGGGCC 353
QY 438 CCACGTCAACAGCGCGGTGCCAGTGCAGAAATGGGGGCTGTGCAACCCCATCACCGGGG 497
DB 354 CCACGTCAACAGCGGTGCCAGTGCAGAAATGGGGGCTGTGCAACCCCATCACCGGGG 413
QY 498 TTGCCATGTGTGCGGGGCTTCCGGGGCTGCGCTGCGAGAGACCGCTGTGAGCAGGGCAC 557
DB 414 CTCGATGTGTGCGCGCGGCTTCCGTGATGGCGCTGCGAGAGACTGTGCGCGCTGGCAC 473
QY 558 CTATGTGAACAGATGTATCAGAGAGTCCAGATGAGAGCCACCTGCGACACACG 617
DB 474 CCACGGCAAGGATGTCCAGCTGTGCGGCTGCGAGTCCAGACAGGTGAGCTGCACCCCG 533
QY 618 CACGGGGAATGCGCGTGCACACAGATATACACCGGAGCTTGTGAGAGATTTGTGCC 677
DB 534 CGCGGGGAGATGCTCTGTGCGACCTGTGCTACACCGCGTCTACTGCGAGAGAGCTGTGCC 593
QY 678 TCTGTGTAACAGTGTCAACAGTGTGAGCAGAGATGCCCTTGTCAAAATGAGAGAGTGTG 737
DB 594 TCTGTGAGAGCAGTGAAGTCACTGATGAGCTGCGCTCCCTGTGCAAAATGAGGAGCAGCTG 653
QY 738 TCATCACTGCTACTGAGAAATGCTCTTCCCTTGTGCTGTGATGGGACAGAGTGTGTGTA 797
DB 654 CCACCATCATCATGCGGAGAGTGTGCTGCGCCCGGCTGAGCGAGGAGAGTGTGTGCCCA 713
QY 798 GCCTTCCCGCGAGAGGTGCGCTTGTGAAAGAACTGTCCCAAGATGCGACAGTGCATTAATG 857

```


QY	1151	AGTGTGCTGC	AACCGGGGCTG	CAAGAACTAC	GTATATG	ACATGTTTCT	CTCGGAT	121
Db	1267	AGTGTCTCTG	CAACCAAGTTTGG	CGGGCTCC	ACTGTCA	CAAGAGCTGC	CTCAGGACA	1328
QY	1211	TCACGGGGAG	CTTGCCACAG	ATGCGAGCT	CGCCAAATGG	GGGCGAGCT	GTGCAGTG	1270
Db	1327	CGCATGGG	CGGGGCTG	CCAGAGACAT	CGCTGTGTG	ACGAGAGGGCT	CTCCCTGGCG	1386
QY	1271	TGACTGGAA	AGTGCACCTGTG	CCCCAGATTC	AAAGAAATG	ACTGTCTAC	CCCATG	1330
Db	1387	ACACGGGCT	CTGTGCGGTG	CGCCCAAGATAC	ACGGGACCT	CACTGCGCTA	CACTATGTC	1446
QY	1331	CTTGGGAA	CCCTATGGGATTA	ACTGTTCCTC	TCGCTGGCTG	GGCTGTA	AAAAATGATG	1390
Db	1447	CACGGGAC	CTTATGGGATTA	CACTGTTCCT	CCCGCTCT	CTGTGAAAAATG	CAATTGCT	1506
QY	1391	GCTCTCC	TGTGAGGGGTCT	TGTACTGTG	GAAGCAGCGCT	GGCAGGGGGT	CACTGTCCA	1450
Db	1507	GCTCTCC	ATTAAGGCGACGTGC	ATCTGCA	AGSAAAGTTGG	CAGCGTGGT	TAATCTGTG	1566
QY	1451	TCAGATGT	CCCAAGTGC	CAATGSGGCT	TGEGCTGT	AACTTAACAT	GCCAGTGC	1510
Db	1567	TTCCCTGT	CCCCCTTGG	GCACACTGGGGGCT	TCAAATGCA	ATGCCAGTGTG	CCACAG	1628
QY	1511	GGGAGCG	CTGCACACCTTG	AGGGACCTG	CAACGTGTG	CACTGGATGG	CGGGGAGA	1570
Db	1627	ATGGAAG	CTGTCAAGCCCCCA	ACAGCTGT	TAATCTTG	CAACCCCTGG	TGGTGTCTC	1686
QY	1571	AATCGA	ACTTCCCTGTG	CAGATGGG	ACAGTACGGGGT	GTAACTGTG	CTGAGCGCTG	1630
Db	1667	ACTCGA	CTTCCCTGTG	CGCCCAAGGAC	AGTTTGTGT	GAAGGCTGTG	CACTGTGTG	1746
QY	1631	GCACGC	ACAGAGATGG	CTGCACACCT	TACAGCGGGC	CAATGCGCTG	CCCTCCGGGATG	1690
Db	1747	GTGCAC	ACTCTG	ATGAGGCTGTG	ACCTGTTC	ATAGGAAGAGCG	CGGATGTG	1806
QY	1631	CAGGTGT	CACTGTG	ACAGCGTGTGTG	GTGAGGAGCG	CTGGGGCCCC	CACTGTCTCTG	1750
Db	1807	TGGGCA	CACCTGTGC	CACTCTCTTGTG	CCGAGGGCTTTT	GTGGGAGCA	CACTGC	1866
QY	1751	CTGTCT	ACTGTAAAAAT	GGGGGCTTC	ATGCTCCCTAT	ATATGAGCA	ATGCGAGTGTG	1810
Db	1867	CTGTACT	GTCAAGAAATG	GTGTGTACTGTGTG	TGTGTGTGT	AGAAATGTG	CAACTGTGTGTG	1928
QY	1811	CAGGCT	TCGAGGAC	CACTTGT	CAGAGATGTG	CTCCCTGGT	TTTATGGG	1870
Db	1927	CAGGTT	TCGAGGGCCCT	CTCTG	CCAGAGGGCC	CGCCGCTGTG	GTGATGTG	1986
QY	1871	GCAC	ACAGACAT	GTCCCACTGTG	CGTTTCA	CAGAGCGGGCC	CTCCACCA	1930
Db	1987	GTGTG	CA-----	ATGCAATGT	TACMACAC	CACTTCTTCTG	CCACCACTG	2040
QY	1931	TGTGT	ACTGTGCT	GTGCTGGCTGT	GCACAGGGCC	CTGTGC	ATATGATGTG	1990
Db	2041	CTGTCT	CTGTGCTGTG	CGGGGCTGTG	AGACAGGGCC	CTGTGCTGT	TAAGGATGTG	2100
QY	1991	GATT	TGGAAAAA	CTGTGTG	CAAGATTTGT	TAACGTGC	CAACAACG	2050
Db	2101	ACTG	GGGAGTCA	ATGTGCTCC	CAACTGTG	CAAGTGTATG	ATGTGTG	2160
QY	2051	TTG	CAACATCT	TGTCACTGT	TATCCCGG	TTTGGATTTG	GCAGTGTG	2110
Db	2161	AGGAT	TGGAGCTGT	ATCTGTG	CAACCGCGCT	GTAGAGCC	CAACTGTG	2220
QY	2111	CAC	CTGGCC	ACTGTGGG	CCAAATGTG	ATCAACAGTGC	ATGCCATATG	2170
Db	2221	CACCA	GAATATG	TTTGTGTCA	CTGTCTCC	AGATGTG	ATGTATGTG	2280
QY	2171	GCAC	CGCTCA	TAGATGGG	GAATGTAA	TGCACTGTG	AGCGAGGGCT	2230
Db	2281	GCAC	CCACAG	ACTGTGGG	CTTGTGTGTG	TTCCCAAG	CAACAGTGTG	2340
Y	2231	AGAG	ATGTC	2239				

D6	2341 TGGGAGGCC 2349
RESULT 6	
LOCUS AF461685	4539 bp mRNA linear ROD 21-JAN-2002
DEFINITION Mus musculus Jcd1-736 protein mRNA, complete cds.	
ACCESSION AF461685	
VERSION AF461685.1 GI:18252657	
KEYWORDS .	
SOURCE Mus musculus.	
ORGANISM Mus musculus.	
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus,	
1 (bases 1 to 4539)	
Krivtsov,A.V., Zinoviyeva,M.V., Hendrikx,J., Vlasser,J.W.M. and Belyavsky,A.V.	
Jedi1 is a novel DSL and EGF-like repeat motif-containing protein expressed on non-differentiated hematopoietic cells	
Unpublished	
2 (bases 1 to 4539)	
Krivtsov,A.V., Zinoviyeva,M.V., Vlasser,J.W.M. and Belyavsky,A.V.	
Direct Submission	
Submitted (20-Dec-2001) Stem Cell Biology, Lindsay F. Kimball Research Institute, New York Blood Center, 310 East 67th Street, New York, NY 10021, USA	
Location/Organisms	
1..4539 /organism="Mus musculus"	
/strain="C57BL"	
/db_xref="taxon:10090"	
/clone="F-4"	
/tisue_type="testis"	
71..2314 /note="soluble form"	
/codon_start=1	
/product="Jedi-736 protein"	
/protein_id="AAI66380.1"	
/db_xref="GI:18252658"	
/translation="MPICPLILALIGRLNGTLNSNDPNVCTEWESPTTTTRESLRP ESLLPASCHRRWEDPHRTCAQPVYTVTYRVKMKSPDLCCRCRYYSRACVP L CADCIVGRCYAFAPMOCOPAEWRSGDSSSEPCAPSMPOCDKFHCNNSSCDPKS G A CFCSGLQPNCQLACPAGAHGPACQDFCCCXYASCPDBDACCPGRARGPSNVN RC SOGDIEFFECPRTPYCQQGVGYOVSQSCLPPGMWGVCISLPEPGHGPCNQDEB RC HNGELDRFTGCQCACAGYLIDRCOECPVGREFGOEATCDCARGARPANACA CL C EHGTGCRTERLCDPGRYGSCSEPCTCPBEHLSCGHPMHGESCOPGNAGLIHNES CPDHPGCGQHNLCKHGGLACLADSLCACBPNTYPHANLPDPTTGINSRS CS ENAIACSPIDGITICEKGMRGCVSPCEDLTGFMCMNSCOUHAHVGSVPQGACT CTPEWHGAHQULPCPKQGFGEBCASVDCDHSDCDVDHQAQRQAOAMMTORCL P PP EGFWGANSTMCITCKNGKTGVSENGMSVCAAFEGFPSCRCPRGARYKRKYCNKN NHSSHSDDSTGSCILAEMTGPDCEAEPFGMHMGTKSCOLCOHHGTGCHPDGSCI TT PGMTGPRCLBGSCPBMRFMVNCSQIQOGDDLDMCHKPIQIACCPRHSHGADCKMESEA PLTRIVEFL"	
CDS	
BASE COUNT 854 a 1399 c 1229 g 1054 t 3 others	
ORIGIN	
Query Match 18.9% Score 648.6; DB 10; Length 4539;	
Best Local Similarity 57.3%; Pred. No. 1.7e-188;	
Matches 1236; Conservative 0; Mismatches 909; Indels 12; Gaps 3;	
OY 74 CTCGTAATCTTAGAACACCCTAATGTGTATGATGTCACCATGGGAAAAGCTTACTGACTGTGC 133 Db 123 CACTCAATCTCCAATGATGCCAATGTCTATTCTTTGGGAAAAGCTTACTGACGACAACTA 182 OY 134 AAAGCTCATACCCACATCCCTTTGATCAA--TTACTACACGAGCTGACTGACATTC 190 Db 183 AGGAGTCCACCTGCGCCCTTGAAGCTGTCCACAGCTGATCTCTGGCCACAGGCCCTGGG 242 OY 191 TTAACGTGTTTAAATGACACGGGGACACAGAATCAGATTCGACACAGCCATAAGACATGGG 250 Db 243 AGGACCCCCACACCTGTGCCAGCCCTACGGTGTCTCTCCAGGACATGTGTACCGTGACAGTGG 302	

251 AGAAGATATATATAGCCCAAGTCTCAGTGTCTGCTGATTTTATGAAACGGGGAAA 310
 303 TGAAGATGAGACTCCCGCCACAGCTGCTGATAGGGGTACTAGAGAGAGAGGG 362
 311 TGTGTGCCCCACATGTGTGATTAATGTGTCATGGTCCGTGATGTATGCTCCAAACCT 370
 363 CCGTGTGCCACTGTGTGCTCCAGAGAGTGTCCATGGTGTGCTGTGTGCTCCGATCAGT 422
 371 GTCAGTGTGAGCCCTGAGTGGGAGAGGACCAATGCTCAGTGTGCTGCGATGTGTATCCT 430
 423 GCCAGTGTGACACAGGCTGGCGGGGTGGCGACTGCTCCAGAGAGTGTGCCCGGAGATGT 482
 431 GGGGTGCCCACTGACACAGCCGGTGGCAGTGTGCAAAAATGGGGCTGTGTGCAACCCATCA 490
 483 GGGGACCAAGTGTGACAAATTTCTGCACATGTGGCAACAAAGTCTGTGATCCCAAGA 542
 491 CCGGGGCTTGCACATGTGTGCGGGCTTCCGGGGCTGGCGCTGGAGAGACCGGTGAGC 550
 543 GTGGGGGCTCTTTTGGCCCTCTGCTGCTGACACCCCAACTGCTTTCAGCTTGGCCCTG 602
 551 AGGGACCTATGTGTACAGCTGTGTATGAGATGCGCAGTGTCCAGATGTGAGAGCCACTGCG 610
 603 CCGGCCACTATGTGTCTGCTGCTGCGCAGTGTGATGCGCAGTGC---TATGGGCACTCTGTG 659
 611 ACCAGTCAAGGGGGGAATGGCGCTGCCACAGATACACCGAGCTTGTGTGAGATC 670
 660 ACCCCAGAGATGGAGCTGTGCTTGTGCTCCAGAGAGAGAGAGAGCCAGCTGTATGTGC 719
 671 TTTGTCTCTGTGTAACATGTGTCCAGTGTGAGAGATGCTTGTGCAAAATGAG 730
 720 CCGTGTACAGGGCACTGATGGGCTTCTTGTGCCCCAGAACCTATCTTGTGCAAAATGAG 779
 731 GATGTGTATCAGCTGTACGTGTGAGAAATGCTTGTGCTTGTGCTGTGTGAGCAAGTGT 790
 780 GTGTCTCTCAGGGCTCCCAAGGCTCCTGACAGTGTGCCACCGGGCTGTATGGGTGCTATTT 839
 791 GTGTGACCTTGTGCCCCGAGGCTGCTTGTGAAAGAACTTGTCCCAAGATGTGCAAGTGC 850
 840 GTTCCCTGCCATGCGCCAGAGGGTTTCCATGTGACCACTTACTCAGAAATGTGCGCTGC 899
 851 ATATAGAGAGGAGCTGTGTATGCTGCAAGAGCCATGTCTATTCAGTGTCCAGATACAG 910
 900 ACAAGGTGGCTCTGTGTGACAGGTTTACTGTGGAGTGGCCACTGTGCTCGGCTATATGC 959
 911 GGGAGGGTGTCCAGATGATGTCTGTGTTGGACCTATGTGCTGTGTGTGAGACT 970
 960 GGGATGGTGTCCAAAGAGTGTGCGGTGGCGCTTGTGCAAGACTGTGTGAGACT 1019
 971 GCCAGTGTGCAAGGAGGAGTGTATACAGTGTGAGCGGCGATGCTGTGAAAGCAG 1030
 1020 GTACTGTGTCTCTGTGGCGCCCTTGTGCTTGTCTATATGCGCGGTGTGTGTGCAACATG 1079
 1031 GCTTGTGTGCGAGCGCTGTGCAAGACGCTGTGTCTGTGAGGGCTGTACAGGATCAAT 1090
 1080 GCTTACAGGGCAGCGCTGTGACTGAGCGCTGTGTGCGGATGTGCGCTGATGTGAGT 1139
 1091 GTATCAAGAGGTGTCTGTGCTGTGCAATGTGAAACACTATAGCTGTCAACCCATGTGTGAG 1150
 1140 GCCAGAGCGCTGTGCTGTGAGCCAGAACACAGTCTCAACCCCTGACAGCGGCG 1199
 1151 AGTGTGCTGCAAGCGGGGTGTGTGAGAGTGTACTGTATGTAGATGTGTCTGTGAT 1210
 1200 AGTGTGCTGTGCGCAGAGTGTGGCGGCTGTCTACTGTCAAGAGAGTGTGCTGTGAGACA 1259
 1211 TCTACGGGAGAGCTGTGCAAGAGATGTGAGCTGTGCAAAAATGGGGAGAGACTGTGAGATG 1270
 1260 CGCATGTGGCGGCTGTGAGAGACTGTGCTGTGTGAGAGAGGAGCTGTGCTGTGCGG 1319
 1271 TGACTGGAAGTGTGACCTGTGCGCCAGAGATGTCAAAAGATGTGCTGTCTACCCCTGCG 1330
 1320 ACAGCGGCTGTGTGCTGTGCGCGCCAGAGATACAGGGAGCTGTCTGCTATACCTATGTC 1379
 1331 CTTGTGGAACTATGGATAACTGTCTGTGCTGTGCTGTGCTGTAAATGTAGTGCAGTGT 1390

1380 CACCGAGACTATATGAGATCACTGTCTCCGCTGCTCTGTGAAATATGCCATGTGCT 1439
 1391 GCTCTCTGTGAGCGGCTTGTACTGTGCAAGGACAGGCTGTGAGTGTGCTGTGCA 1450
 1440 GCTCTCCATGTAGAGGACAGTGTGATGTGCAAGAGAGGTTGAGCGGTGTATGTGCTGTG 1499
 1451 TCAGATGTCCAGTGTGACATGTGGGCTTGTGCTGTATTAATATGACAGTGTGCTCAAG 1510
 1500 TTCCCTGTCCCTTGTGAGACTGTGGGCTTGTCAATGTGCAATGTGCAATGTGTGCTGAG 1559
 1511 GGGAGCTGTCAACACCTGTGAGCGGAGCTGTGACGTGTGCACTGTGATGTGCGGAGAGA 1570
 1560 ATGAGAGTGTGACACCCCAAACTGTGAGAGCTGTACTGTGACACCCCTGTGATGTGCTG 1619
 1571 AATGCGAATCTCTGTGCAAGAGTGTGAGAGTGTGAGAGTGTGAGTGTGCTGTGAGCTGTG 1630
 1620 ACTGCGAGCTTCCCTGTGCGCAAGGAGAGTGTGTGAGAGTGTGAGTGTGTGTGCTGTGACT 1679
 1631 GCAGCCAGCAGATGTGCTGTGCAACAGGGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1690
 1680 GTGACCACTGTGATGTGTGTGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 1739
 1691 CAGTGTGTGCTGTGACAGCTG 1750
 1740 TGGGCAACAGCTGTGCAACCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1799
 1751 CTTGTACTGTGTAATATGAGGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1810
 1800 CTTGTACTGTGTAATATGAGTG 1859
 1811 CAGCTTGTGAGGACCACTGTGTGAGAGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 1870
 1860 CAGGTTGTGAGGACCCCTG 1919
 1871 GCAGCAGAGATGTGCAAGTG 1930
 1920 GTGTGCA-----ATGCAAGTG 1973
 1931 TGTGTACTG 1990
 1974 CTTG 2033
 1991 GATTGTGGAAGAACTG 2050
 2034 ACTG 2093
 2051 TTGACAGATG 2110
 2094 AGGATG 2153
 2111 CACCTGCCCCAGTG 2170
 2154 CACCAAGATG 2213
 2171 GCAGCGCTG 2227
 2214 GCCACCCAGAGATG 2270

RESULT 7
 AR440279
 LOCUS AF440279 4482 bp mRNA linear ROD 20-NOV-2001
 DEFINITION Mus musculus MEGF12 (Megf12) mRNA, complete cds.
 ACCESSION AF440279
 VERSION AF440279.1 GI:17017250
 KEYWORDS
 SOURCE Mus musculus.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 4482)
 AUTHORS Ivanova, N.B. and Lemischka, I.R.

1511 GGGGAGCCCTGCAACACCTTGAGAGGAGCTGCACTGTGACCTGTGATGGCGGGAGAG 1570
 1827 ACGAGACTGTGACAGCCCCCAAACTGGAGCCTGTGATGCAACCCCTGGGTGGCATGTGCTC 1886
 1571 AATGCAACTTCCCTGCGCAGAGATGGCAGCTAGCGGGCTGAACTGTGCTGACCGTGGACT 1630
 1887 ACTGCCAGGTTCCTCCCTCCGAAAGGAGAGTGTGTAAGAGCTGTGCTGCTGTGACT 1946
 1631 GCAGCCAGCAGATGGCTGCGACCTTACCAAGGGCCATTGGCCGCTCCCGGGATGGG 1690
 1947 GTGACCACTGTGATGGCTGTGACCTGTTCATGAGACAGTGGCGATGTCAAGCTGGTTGA 2006
 1691 CAGGTGTCCAGCTGTGACAGCAGCTGTGCTGAGGAGCAGTGGGGCCCAACTGCTCCCTGC 1750
 2007 TGGGACACAGCTGCCACCTGCTGCGGAGGGCTTTTGGGAGGCACTGCACTGCACTAACA 2066
 1751 CCGTCTACTGTATAAATGGGGCTTCACTGCTCCCTGATGATGAGCACTGTGAGTGTGAC 1810
 2067 CCGTCTACTGTCAAGAAATGGTGTGCTGTGTGTGAGAAATGGCACTGTGCTGTGCTGAC 2126
 1811 CAGGCTTCCGAGGACACCTGCTGACAGAGATGCTCCCTGCTTTTATGGGATGGT 1870
 2127 CAGGTTCCGAGGACCCCTCTGCTGCAAGAGCCCTGCGCTGCTGCTATGCAAAAGCT 2186
 1871 GCAGCCAGACATGCGCCACAGTGGCTTACAGAGCAGCGGGCCCTGCCAATCACCAGGCC 1930
 2187 GTGTGCA-----ATGCAAGTGTAAACACACCATTTCTGCTGCAACCATCGGAGGGGA 2240
 1931 TGTGTACTGTCTGCTGCTGCTTACAGAGGCGCTCTGCAATGAAATGTGTCCCAAGTGGCA 1990
 2241 CCGTCTCTGCTGCGGGCTGTGACAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2300
 1991 GATTGTGAGAAACTGTGACAGAAATTTGATCTGACACCAACAGGAACTGTAAACCCA 2050
 2301 ACTGGGAGCTCAATCTCTCCAACTGTGCAAGTGTATGATGGTGGGAGCTGCAACCC 2260
 2051 TTGACAGATCTGTGATGCTTACCCGCTGTTGATGGATGGAGTGTGCTTCAACATGTC 2110
 2361 AGGATGGAGCTGTATCTGACAGCGCCAGGCTGAGTGGACCACTGCTGGAAGGCTGCC 2420
 2111 CAGCTCCGACAGTGGGCGCAAACTGCATCCACAGCTGCAACCTCCATATGAGCTTCT 2170
 2421 CACCAAGAAATGTTGTGTCAACTGCTCCAGCTATGTGATGTGATGCGGAGATGT 2480
 2171 GCAGCCCTACAGATGGGAGATGTAATGCACTCTGCTGAGAGGGCTTACTGACATC 2230
 2481 GCACCCAGAGACTGGGGCTTGT 2540
 2231 AGAGATGTC 2239
 2541 TGGGAAGCC 2549

RESULT 8
RATORD 660 bp mRNA linear ROD 09-AUG-1996
LOCUS Rattus norvegicus (clone REM4) ORF mRNA, partial cds.
DEFINITION L41686
VERSION L41686.1 GI:780366
KEYWORDS monoclonal autoantibody;
SOURCE Rattus norvegicus (strain Holzman) cDNA to mRNA.
ORGANISM Rattus norvegicus
REFERENCE 1 (bases 1 to 660)
AUTHORS Asakura, K., Pogulis, R.J., Pease, L.R. and Rodríguez, M.
TITLE A monoclonal autoantibody which promotes central nervous system remyelination is highly polyreactive to multiple known and novel antigens
JOURNAL J. Neuroimmunol. 65 (1), 11-19 (1996)
MEDLINE 96235155

PUBMED 8642059
FEATURES
source
 location/Qualifiers
 1..660
 /organism="Rattus norvegicus"
 /strain="Holzman"
 /db_xref="taxon:10116"
 /clone="REM4"
 /issue="type="Brain"
 /dev_stage="neonate"
 <1..>660
 <1..>660
 /note="ORF"
 /codon_start=1
 /protein_id="AAB05844.1"
 /db_xref="GI:780367"
 /translation="RGATCDHTTSECRSPGYTGAFCEDLCPRGKHGPOCRPCON
 GVCCHHTVEGSCPSGMMGYGQPCPSRBRKNSQSCCHENAGCAATATGQCHSP
 GTTGERCODECVGVVCAETCRVNGKACVHVSCTLCESGSEFCARLCPG
 LYGINCDRCPCHLDNTHSCHPMSECCCKPMSGLYCNETCSPPGYBACOOICSQ
 NG"

BASE COUNT 141 a 160 c 205 g 154 t
ORIGIN
 Query Match 16.0%; Score 548.2; DB 10; Length 660;
 Best Local Similarity 89.6%; Pred. No. 1.4e-157;
 Matches 589; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

598 GAGGACACCTGGGACACGACGTCACAGGAGATGCGCTGCGCCACAGATACACCGAGCC 657
 4 GAGGACACCTGTGACCAATCACTGAGGAAATGCTGTTCACTGATACACTGAGCC 63
 658 TTCTGTGAGATCTTGTCTCTCTGTGTAACATGATGTCACAGTGTGAGCAGAGATCCCT 717
 64 TTCTGTGAGATCTTGTCTCTCTCTGTGTAACATGATGTCACAGTGTGAGCAGAGATCCCT 123
 718 TGTCAAAATGAGAGATGTGTATCATCATGCTGAGGAAATGCTCTTGGCTTGTGCTGG 777
 124 TGTCAAAATGAGAGATGTGTATCATCATGCTGAGGAAATGCTCTTGGCTTGTGCTGG 183
 778 ATGGGACAGTGTGTGTGAGCTTGTGAGGAGGCTGCTTGTGGAAGAACTGTTCCAA 837
 184 ATGGGACAGTGTGTGTGAGCTTGTGAGGAGGCTGCTTGTGGAAGAACTGTTCCAA 243
 838 GAATGCCAGTGTGTGTGAGGAGGCTGTGATGCTGTCACAGGCAATGTCAATGCACT 897
 244 GAATGCCAGTGTGTGTGAGGAGGCTGTGATGCTGTCACAGGCAATGTCAATGCACT 303
 898 CCAAGATACAGAGGAGGAGGCTGTGATGCTGTGAGGAGGCTGTGAGGAGGCTGTGAG 957
 304 CCAAGATACAGAGGAGGAGGCTGTGATGCTGTGAGGAGGCTGTGAGGAGGCTGTGAG 363
 958 TGTGTGAGACCTGCGCAGTGTGTCAACAGGAGGAGGAGGCTGTGAGGAGGAGGCTGTG 1017
 364 TGTGTGAGACCTGCGCAGTGTGTGTATGAGGAGGAGGAGGCTGTGAGGAGGAGGCTGTG 423
 1018 CTCTGTGAAGCAGGCTTCTGTGCGAGCGCTGTGGAAGACGCGCTGTGCTGAGGGCTC 1077
 424 CTATGCGAAGCAGGCTTCTGTGCGAATTTGTGAGGAGGCTGTGCTGCGCTGAGGGCTT 483
 1078 TACGGCATCAATGTGACAAAGGCTGCTGCGCACTGTGGAAGGAGGCTGTGAGGAGGCT 1137
 484 TACGGCATCAATGTGACAAAGGCTGCTGCGCACTGTGGAAGGAGGCTGTGAGGAGGCT 543
 1138 CCATGTGTGAGAGTGTGCTGCAAGCGGCTGTGAGAGTGTGAGAGTGTGAGAGTGTGAG 1197
 544 CCATGTGTGAGAGTGTGCTGCAAGCGGCTGTGAGAGTGTGAGAGTGTGAGAGTGTGAG 603
 1198 TGTGTCTGTGATTTACGGGAGGAGGCTGTGCAAGCAGATGTGAGCTGCCAAATATGGG 1254
 604 TGTGTCTGTGATTTACGGGAGGAGGCTGTGCAAGCAGATGTGAGCTGCCAAATATGGG 660

RESULT 9

AX079681
 LOCUS AX079681 632 bp DNA Linear PAT 22-FEB-2001
 DEFINITION Sequence 425 from Patent WO0107611.
 ACCESSION AX079681
 VERSION AX079681.1 GI:13159250
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 1 (bases 1 to 632)
 AUTHORS Baker, K.P., Goddard, A. and Wood, W.I.
 TITLE Human polypeptides and methods for the use thereof
 JOURNAL Patent: WO 0107611-A 425 01-FEB-2001;
 Genentech, Inc. (US)
 FEATURES
 Location/Qualifiers
 source 1..632
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 BASE COUNT 145 a 159 c 172 g 154 t 2 others
 ORIGIN
 Query Match 15.6%; Score 534.8; DB 6; Length 632;
 Best Local Similarity 95.3%; Pred. No. 2e-153;
 Matches 593; Conservative 0; Mismatches 24; Indels 5; Gaps 4;

LOCUS AK074121 4470 bp mRNA Linear PRI 15-FEB-2002
 DEFINITION Homo sapiens mRNA for FLJ00193 protein.
 ACCESSION AK074121
 VERSION AK074121.1 GI:18676589
 KEYWORDS fls (full insert sequence).
 SOURCE Homo sapiens adult spleen cDNA to mRNA, clone:FLJ00193.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 1 Jikuya, H., Takano, J., Nomura, N., Kikuno, R., Nagase, T. and Ohara, O.
 TITLE The nucleotide sequence of a long cDNA clone isolated from human spleen
 JOURNAL Published Only in Database (2002)
 REFERENCE 2 (bases 1 to 4470)
 AUTHORS Jikuya, H., Takano, J., Nomura, N., Kikuno, R., Nagase, T. and Ohara, O.
 TITLE Direct Submission
 JOURNAL Submitted (21-JAN-2002) Takahiro Nagase, Kazusa DNA Research Institute, Department of Human Gene Research, 1533-3, Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:cdnainfo@kazusa.or.jp, URL: http://www.kazusa.or.jp/NEDO, Tel:81-438-52-3913, Fax:81-438-52-3914)
 COMMENT NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert and 5'- & 3'-end one pass sequencing; Research Association for Biotechnology; cDNA library construction and clone selection; Kazusa DNA Research Institute.
 FEATURES
 Location/Qualifiers
 source 1..4470
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="FLJ00193"
 /tissue_type="spleen"
 /dev_stage="adult"
 /note="vector: pBluescriptII SK plus"
 1..4470
 /gene="FLJ00193"
 <1367..2842
 /gene="FLJ00193"
 /note="Start codon is not identified."
 /codon_start=1
 /evidence="not_experimental"
 /product="FLJ00193 protein"
 /protein_id="BAB84947.1"
 /db_xref="GI:18676590"
 /translation="PCSWTLVDFGFMGKARHLSPEELKGVNCSNQTCKNGSTCLP
 ENGNVCAPGRGSPSCQSRGGRKRCVPCCKANHSFCHSNGTCYCLAGWGPDC
 SQPPRGHMGCAOTCCQNHGTCCHPDGSCICPLGWTGHHCEGLGTFGAGNSQ
 PCQSGPKCKPRTGACVCPRHSGAPRIGIOEPTVMPPTPAVNLGAVIGAVL
 GSVLVALALFITYRHMKQKEHNHLLAVASSGRLDGSEVYMPVPPVSHYSNPSY
 HTLSQSPNPPRPKVRGRLDASLQNPERRPGAGQHDHNTLLPMDKRRRPPRPGLD
 RGSRLDRSYSTYSNGPFRYKGLISEELGASVASLSENPATIRDLPSLPGGP
 RESSYEMKGRPPSSPPRPQPFMDSORRPPQPRQSGTVEQPSPLIHDDSDVSGSP
 PLPLPLPGHDSKRNSHIPGHYDLPPVRHNPSPPLRRQDR"
 BASE COUNT 888 a 1372 c 1240 g 970 t
 ORIGIN
 Query Match 13.3%; Score 456.8; DB 9; Length 4470;
 Best Local Similarity 55.5%; Pred. No. 4.4e-129;
 Matches 1067; Conservative 0; Mismatches 767; Indels 90; Gaps 6;

QY	541	CGCTGTGACGAGGGGACCACTTGTGTAAAGACTGTGATCATAGAGATGCCAGTGGCCAGTAATGGA	600
Db	181	CCCTGTACCCCTGGCTACTAATGAGCCCTGCCTGCCAAGTTCGCGTCCAGTGTCC---ATGGG	237
QY	601	GCCACCTTCGACCAACGTTCACGAGGGGAATGCCCGTGCCTCCACACAGATATCACCCGAGCCTTTC	660
Db	238	GCACCCCTCCGATGCCAGACTGTGGAGCCCTGCTTGTGCCCGCCAGAGAACTGGGCCACAC	297
QY	661	TGTGAGATCTTCTTCCCTCCCGGTAAACATGATGATCCACAGTGTATGAGAGATATGCCCTT-G	719
Db	298	TGTGACGTGTCTCTTTCCTCCAGGGCACTTTCGCTTCTTCTTGCCCCAGACCCCACTCCCTGG	357
QY	720	TCAAAATGAGAGAGTGTGTATCATCAGTCACTGGAGATGCTCTTGGCTTTCGGCTGTGAT	779
Db	358	CCAAAATGAGAGTGTCTTCCAAAACCCACAGGGGCTCCTGAGCTGTGCCCTCGGGGTGAT	417
QY	780	GGGCAACGTGTGTGTGAGGCTTGGCCCCGAGGGGTGCGCTTTGGAAAATACTGTTCCTCCAGA	839
Db	418	GGGCACCATCTGTCTCCCTGCTGCCCTCCAGAGGGGCTTTACAGGACCCCACTGTCTCCAGGA	477
QY	840	ATGCCAGTGCCTAATATGAGAGGAGCAGTGTGATGCTGCCACAGGCAAGCCAACTGTATTCAGTCC	899
Db	478	ATGTGCTGTCCACAAACGGGCGCCTGTGTACCATTCACGTGGGCAATGCGGTGGGCTCC	537
QY	900	AGGATACACAGGGGAAC-----	916
Db	538	GGGTACACTGGGGATGCTGTGAGTGGCGTGGGCGGGGCGGAGACGGAGAGAGAGTGTG	597
QY	917	-----GTTGACAGAGATGATGTCTCTGTGTTGGGA	943
Db	598	GGGGGGCGGACAGGCTCACTACGTAAGTGAAGTGGCCAGAGTGTCCGGAGAGATGCTCCGGTGGCC	657
QY	944	CCTATGCGCTTCTCTGTGCTGTGAACCTGCCAGTGTGTCAACGAGAGGGAAGTTATCACAG	1003
Db	658	GCTTTGGGACAGAACTGTGCTGTGAACAGCTGACCTGCCCGGAGCGCCCTGTCTTCCGG	717
QY	1004	TGACGGCGCATGTGCTGTGTGAAGCAGGCTTGTCTGGGAGAGGCTGCGGAAGCAAGCAAGCCGT	1066
Db	718	CCAAAGGCGCATGTCTGTGCGAACACGGCTTCACTGGGGAACGGCTGCACGATGCGCTCT	777
QY	1064	GTCCTGAGGGGCTCTACGGCATCAAAATGTGTACAAACGATGTCCTCTCACTTGGAAAACA	11223
Db	778	GCCCCGACGGCTTCTACGGTCTACAGCTGACGACGGCCCGCTGCACCTGCGACGGAGGACMA	837
QY	1124	CTCATACTCTACACCCCATATCTGTGAGAGAGTGTGCTTCGCAAGCCGGGCTGTGAGACATCT	1188
Db	838	GCTCTACCTGTCCACCCGATATACGGGAGAGTGTCTCTGCTCCCGGGCTGTGGGGGCTTCC	897
QY	1184	ACTGTAAATGAGACATGTTTCTCTGTGATTTCTACGGGGAACCTTGTCCAGCAGATCTCCACT	1243
Db	898	ACTGCACGAGAGAGCTGTGCCGAGAGACAGGATGGGCGAAGGTGTCCAGAGCACTGTCTCT	957
QY	1244	GCCAAATGGGGCGAGACTGTGACAGTGTGACTGTGAGTGTGAAATGTGACACTGTGTGCCAGATTTCA	1303
Db	958	GCTGTGACGGTGTGCTGTGTGACAGGCTTACAGAGGGCCCTGTGATATGTGCGGGCGGGTGTACA	10174
QY	1304	AAGGAATGTGACTCTCTACCCCAATGCTGCTGTGGGAACATATGGGATTAACCTTCTCTCTG	1366
Db	1018	CGGGCCCTCACTGTGTGATGCTTGTGCTCTGTGACACCTACAGGTGTCAACTTCTTGTGAC	10777
QY	1364	GCTGTGGCTTAATAAATGATGACGTGTGCTCTTCTGTGTGAGCGGGCTGTGTACTTGCAGAG	14223
Db	1078	GCTGTCTCATGTAAATATGCCATGGCTGTGTCAACCCATCGACGGGCAAGTGTGCTTCAGAG	11373
QY	1424	CAGGCTGTGACGGGGGTGAGCTGTGCATCAGATGTGCCAGTGTGCCACATGTGGGCTTGGCT	14833
Db	1138	AAGGTTTGGCAGCTGTGTAATGTCTGTCTGTCCCTGTCCACCCGGAACCTGTGGGCTTTCAGTT	1197
QY	1484	GTAATTAACATGTCCAGTGTCTTCAAGGGGAGAGCTGTCAACACCTGTGACGGGAGCTGTCA	15433
Db	1198	GCAATGCGACACTGCCAGTGTGTGCCCATGTAGAGGAGTGTGTGAGAGCCCAAACTGTGAGCTTGA	1257
QY	1544	CGTGTGACCTGGATGTGGCGGGGAGAAATGTGCAACTTCTGTCAGGATGTGACGTATCG	1603

Db	1258	CTTGCACCCCTTGGGTGGCATGTGGGGCCCATCTGCCAGCTTGCCTGTCCGAAGGGGCAATTGG	1317
QY	1604	GGCTGAACCTGTGCTGAGCGCTGCGATGCGAGCCAGCAGATGGCTGCCACCCATCCACG	1663
Db	1318	GAGAAAGTGTGTCCAGTGGCTGTGACTGTGACCACTCTGATGGTGTGACCCGTTCATG	1377
QY	1664	GCCATTGCCCGCTGCCCTCCCGGATGTGTACAGTGTGCCACTGTGACAGCGTGTGTGAG	1723
Db	1378	GACGCTGTCAAGTGCAGAGG - TTTCTGGATGGGTCCCGCTGCCACCTGTCTGCCCTGAG	1436
QY	1724	GACGCTGGGGCCCCAAGCTGCTCCGTCGCTGTACTGTAAATAGGGGCTTATGCTGCC	1783
Db	1437	GCTTATGGGGAGTCAACTGTATGCAACACCTGTGACCTGCAAGATATGGGGCACCTGTCTCC	1496
QY	1784	CTGATGATGGCACTCTGCGAGTGTGCACACAGGCTTCCGAGGCAACATTTGTCAAGAGATCT	1843
Db	1497	CTGAATAATGGCAACTGCGTGTGTGCACCCGGATTCGGGGGCCCTTCGCAAGATCGT	1556
QY	1844	GCTCCCGTGGTTTTATGGGCAATCGCTGCAGGCACAGCATGCCACAGTCCGTTACAGAGA	1903
Db	1557	GTCAACCTGGCGCGCTATGGCAAAACCTGTGTGC - - - - - CTGCAAGTCCGGTTAACCACT	1610
QY	1904	GCGGGCCCTGTGCCACACATCACCGGCTGTGTGACTGTTGGCTGGCTTCACAGGGGCC	1963
Db	1611	CC - - - TTTGTGCCACCCCTCGAAGCGGAGACTGCTACTGCTCTGGCTGTGACAGGGCCGG	1667
QY	1964	TCTGCAATGAAGTGTGTCCCACTGGCAGATTTTGGAAAACTGTGCAGGAATTTGTACT	2023
Db	1668	ACTGCTCCAGCACTGCCCTCCAGAGACACTGGGGGAAAACTGTGCCAGACTGTCCAMT	1727
QY	2024	GCACCAACAAGGAGACCTGTAAACCCCATTTGACAGATCTTGTAGTGTATACCCGGTTGGA	2083
Db	1728	GTCAACCAATGTGTGGGACCTGCCATCCCAAGATGGAGCTGTATCTGCCCCCTAAGGCTTGA	1787
QY	2084	TTGGCAGTAGACTGTCTCAACCATGTCCACACTGGGAGCCCAAACTGCATCCACA	2143
Db	1788	CTGGACACCACTGCTTTAGAAAGGCTCCCTCTGTGGGACATTTGTGTCTAACTGCTCCACG	1847
QY	2144	CGTGCAACTGCCATATATGAGGCTTTCTGCAGCGCCTACGATGGGGATATTAATGCATCT	2203
Db	1848	CATGCGACTGTGTGTCCTGTGAGAAAGTGCACCCAGAGACTGGGGCCTGTATGTGCC	1907
QY	2204	CTGG 2207	
Db	1908	CAGG 1911	

RESULT 11	AB011532				
LOCUS	AB011532	5523 bp	mRNA	linear	ROD 22-AUG-1998
DEFINITION	Rattus norvegicus mRNA for MEGF6, complete cds.				
ACCESSION	AB011532				
VERSION	AB011532.1	GI:3449293			
KEYWORDS	MEGF6.				
SOURCE	Rattus norvegicus (strain:Sprague-Dawley) adult brain cDNA to mRNA, clone:11b:psFOR 1 clone:RG1805.				
ORGANISM	Rattus norvegicus				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.				
REFERENCE	1 (sites)				
AUTHORS	Nakayama,M., Nakajima,D., Nagase,T., Nomura,N., Seki,N. and Ohara,O.				
TITLE	Identification of high-molecular-weight proteins with multiple EGF-like motifs by motif-trap screening				
JOURNAL	Genomics 51 (1), 27-34 (1998)				
MEDLINE	98360089				
REFERENCE	2 (bases 1 to 5523)				
AUTHORS	Nakayama,M., Nakajima,D. and Ohara,O.				
TITLE	Direct Submission				
JOURNAL	Submitted (26-FEB-1998) Manabu Nakayama, Kazusa DNA Research Institute, Laboratory of DNA technology; 1532-3, Yana, Kisarazu,				

Chiba 292-0812, Japan (E-mail: mmanabu@kazusa.or.jp,
Tel:+81-438-52-3915, Fax:+81-438-52-3914)

FEATURES

Source

Location/Qualifiers
1..5523

/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="RG1805"
/tissue_type="brain"
/clone_1lb="psport 1"
/dev_stage="adult"

gene

1..5523

CDS

/gene="MEGF6"
235..4959
/gene="MEGF6"
/codon_start=1
/product="MEGF6"
/protein_id="BA32462.1"
/translation="MPVRAEAAARVVALALLLPAMPASPLTPRIPIPSMPHYC
ABOKLTIHROPQAFRIYPIVPRIRGCAQQAACIGQERRTYVYMYROYATKAR
TVRRCPGRSOKPQGECLSDYDECASANGCEBPCNTVGFTICRPPGIDQDGK
TCQDYDECAHNGCQHRVNTPSYLCEKRGFLHTDRTCLAISSTLGNCGCH
QCVQLTVQHRCQRPYQLQEDGRVRRSPCAENGCMHICQELRGLAHCCHPG
YLAADRKTCEVDICALGLAQCAGHCNTOGSPFCVCHAGYELGADGRORYMEI
VNSCEAGNGSGSHGSHSTGTPLCTCPRGYELDEDKTICIDIDCANSPPCCQAANT
PGYECSCFAGRLNTDGGCEDVDECAHSGHSCCHHNSIANGSPCCCEAGRYED
RRCSTLESVYVDLGRIPVPRPLPHIYVLRDELRLRDDYGAEEBAABAELEHNT
LTERKVCIDHSFGHDCSLTDCDRNGGTCFPOQDCCDEPEGTGILICETCPDIFGK
NCSSPCTQNGGTCDPVLGACRCPPVGAHCEDECPKRGKHKRKHCHCNRGKH
RLYGALCDPGLYGRFCHLACPPAFGPGSCEDICEQSHTSPCKDKSCSKAGFO
GERQCAEESGFFGPGRRHRCQCPGVACDPVSGSCPTQCPRGYEGEOGOCGPGTE
GNSGSCSGVAGPCRHVYRTECLCPRGTEGEGADCEBGRMKGLCCOETLCPACSEAS
CNPETGTCLCPGFGSRCDPTCSNWRGTGTGQITKACANDGHCPTTGRSCAAGWT
GLSCORACDSGHWGPDCTHPCNCSAGHNCDSVSLCLCEADYEBPBEQSCRGTYG
PSEQRKCRERHAACDVAHGAATCPAGMRGSPCEHACPAFEGDLDSCNCSAAPC
DAVYESICIPAGRMGPRCAQSCPLPFLGLNSLOICTENGASCDSDVTCQCHAPMG
PTLOACPPGLYKNCQSHCLCRNGRCDPLIGQCTCEPMTGLACENCLRGHYAAG
COLNCSLHGGLCDRLTGLHCLCPAGMTDKOOSCSVSTFGVHCEHCAKRGASCH
VTGACFCPPMRGPHCEBACPRMRGEACQRCLEPTNASHVHYEBRCRPPGTGLS
CBQACQPGTFCBDCBHLQCPETWACDPASGVCTCAAGYHGTGLQCRPSRYTPGC
EHICCLNGGTCDDPATGACYPAGFLGADCSLACPOGRFPGCAHVCACROGAADPV
SGACICSPKGTGVRCEHGCPODRFEGCEELKACNGLLCHATNNGSCPLGMMPHC
BHACPARGYAACLCEFCQNNNGSC EPTTGACILGPRGYGACCHESGPGFPCOR
VCEQOQGAAPDPVSGOCLCPAGFHGRCPEKGECSRFDGCLQCCNCHGVYCDPISG
LCLECPRTGACACDLCRRRGRPECALRCDCGGGADCDPISGQCHVDSTWGTGRE
VPTQISSRPAVQHPHSRAKMH"

polyA_site

5523

BASE COUNT 960 a 1603 c 1797 g 1163 t
ORIGIN

Query Match 11.7%; Score 400.8; DB 10; Length 5523;
Best Local Similarity 51.3%; Pred. No. 1e-111;
Matches 1092; Conservative 0; Mismatches 1017; Indels 21; Gaps 6;

QY 372 TCAGTGTGAGCTGCTGGGGAGGAGACCAACTGCTCCAGTGCCTGGGATGGTGTGATCAGT 431
DB 2394 TCAGTGTGAGCTGCTGGGGAGGAGACCAACTGCTCCAGTGCCTGGGATGGTGTGATCAGT 2453
QY 432 GGGTCCCACTGACACGACCGGTGACCAAAATGGGGCTGTGTGACACCCCATCAG 491
DB 2454 TGGTGTGAACTGCTCAGGTCTCTGTTCTCTGC---GTTGAGAGCGCTTGTACCGAATCAG 2510
QY 492 CGGGGCTTGCACCTGTGTCTGGGGGCTTCGGGGCTGGGCTGTGGAGAGCCGCTGTGAGA 551
DB 2511 AGGGGAGTGTGTGTGCGCCACAGGAAAGACTGGGGAGAGACTGTGAGTGTGTCGCA 2570
QY 552 GGGCACCTATGTGAAGCATGTGCATCAGAGATGCC---AGTGCACAATAGGAGCCACTG 608
DB 2571 GGGCGCTTGGGGGCTGGGTCTCAAGAGATGTGCCCGGACATGTGACACGAGACCGACGCTG 2630

QY 609 CGACACGATCACGGGGGAATGCCGCTGCCACACGAGTACACCGAGCCTGTGTGAGA 668
DB 2631 CAACCTTGAGACTGGAACTGCTGTGTCTCCTGTGCTTGTGTGAGCCGCTGCCAGGA 2690
QY 669 TCTTTCTCTCTGTGTAACATGCTGACAGTGTGACAGAGATGCTGTGCAAAATG 728
DB 2691 CACTTCTCAGAGCGCTGTGACGGAGCTGGCTGCCAGATTAAGGTGTGCTTGTGCCAAGA 2750
QY 729 AGGAGTGTATCATCAGTCACTGAGAGATGCTTGTCCCTTCTGTGCTGTGATGGGCACAGT 788
DB 2751 TGGGCAATGTGTACCCACCAACGAGCGCTGTACTGTGCCCCGGGTGTGACTGCGAG 2810
QY 789 GTGTGTGACCTTGTGCCCGGAGGCTGTTGGAAAGAACTGTTCCAAGAAATGCCAGTG 848
DB 2811 CTGCCAGAGAGCTGTGACAGTGTGACACTGGGGGCCCGACATGATCACCCTTGCACACTG 2870
QY 849 CCATATATG---AGGACGCTGTGATGCTGCCACAGGCCAATGTCAATGTGACAGATTA 905
DB 2871 CAGTGGGGCCATAGGAATTTGTACGCCCGTCACTGCGCTGTGTGAGGCCGCGCTA 2930
QY 906 CACAGGGGAACGGTGCACAGATGATGTCTGTGTGGACCTATAGCGCTTCTGTGTCTGA 965
DB 2931 TGAAGGTCCACGCTGTGACACAGTGTGTGCCAGAGGATACTATGAGGCCACAGCTGAGGA 2990
QY 966 GACCTGCCAGTGTGTCAACGAGGAGGAAGTGTACACAGTGAAGGGGGCATAGCCTGTGTA 1025
DB 2991 GAAGTGCATATGTGACATATGGGCTCCTGTGACATGTACAGGGGGCTGTATACCTGCC 3050
QY 1026 AGCAGGCTTGTCTGGGAGCGCTGCGAGACAGCAGCCTGTGTCTGTGAGGGGCTTACGGCAT 1085
DB 3051 AGCGGCTGTGAGGGGAGGAGCTGTGTGTGAAATGCG---GTGCGCTGTGCTTGTGGT 3107
QY 1086 CAAATGTGCAAAAGGTGTCCCTGCCACTGTGAAAACACTCAATAGTGTACCCCATGTG 1145
DB 3108 AGCTGTGTGACA-----GTGCTGTGCAACGTCTGTGCGGGGACCTGTGTGATGTGTGAC 3161
QY 1146 TGGAGAGTGTGCTGTGCAACCGGGCTGTGACAGACTCTACTGTAATGAACATGTTCTGC 1205
DB 3162 CGGCTCATCATGTGTGCCAGCTGTGCGGCTGTGGGCCACACCTGTGGCCACAGTTGCCAC 3221
QY 1206 TGGATTTCAAGGGGAAGCTTGTGCCAGAGATCTGACACTGCCAATATGGGGGACACTGTGA 1265
DB 3222 CTTACACTTCTGGGCTTAACCTGACAGCAGATTTGTACTGTTCATATGGGGGCTTCGTGA 3281
QY 1266 CAGTGTGATGGAATGTGACACTGTGTGCCAGAGTTCAAAGGAATATGATGCTCTACCC 1325
DB 3282 CTCTGTCAAGGGCAATGTGCACACTGTGCCCTGTGATGGAGCCACACTGTGTCGAGGC 3341
QY 1326 ATGCCCTGTGGGAACCTATGGATTAAGTGTCTCTGCTGTGGCTGTAAATATGATGC 1385
DB 3342 CTGCCCTCTGGCTGTATGTGCAAGAACTGTGACATTTCTGTCTTGTGCCGCAAGGAG 3401
QY 1386 AGTGTCTCTCTGTGAGAGCGGTCTGTACTTGTCAAGGAGAGCTGTGACGGGTGTGAGT 1445
DB 3402 GAGATGTGACCCCAATCTTGGGCGACAGTGCACAGAGGTTGAGACCGGCTTGGCCG 3461
QY 1446 CTTCATCATATGTCCAGTGTGACATATGGGCTTGTGCTTAATTAATGCCAGTGCCT 1505
DB 3462 TGAAGAACGAGTGTCTGTGACACTATGACGCGGTGTGTACACTCAACTCAAGTTGCTT 3521
QY 1506 CAAGGGGAGGAGCTGTGACACCTGTGAGAGGAGACTCTGACAGTGTGACAGTGGAGCGG 1565
DB 3522 TCAGGGGGGAGATATGTGACCGGCTTACCGGCACTGTCTGTGCCACAGCTGTGAGACTG 3581
QY 1566 GGAAGAAATGCAACTTCTGTGCCAGAGTGTGACAGTGAAGGCTGTGACTGTGAGCGCTG 1625
DB 3582 GGAACAAGTGTCAAGAGCTCTGTGCTGAGTGTGAGCTTGTGGGGTGTCACTGTGAGAAACAGT 3641
QY 1626 CGACTGTCAACCAAGCAGATGAGCTGTCCACCTTACCAAGGGCCATATGCGGCTGTCCCGGG 1685
DB 3642 TGTGTGCGGAAAGGTGCTCTGTCCACCATGTGACAGGGGCTGTGCTGTGTCTCAAG 3701
QY 1686 ATGTGTGAGTGTTCACACTGTGACAGCGTGTGTGTGAGGAGACCTGTGGGCGCCCAACTGTCT 1745

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 146206)
AUTHORS DOE Joint Genome Institute.
JOURNAL Direct Submission
TITLE Submitted (24-MAR-2000) Production Sequencing Facility, DOE Joint
AUTHORS Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 3 (bases 1 to 146206)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
JOURNAL Direct Submission
TITLE Submitted (28-MAR-2001) DOE Joint Genome Institute, 2800 Mitchell
AUTHORS Drive, Walnut Creek, CA 94598, USA
REFERENCE 4 (bases 1 to 146206)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
JOURNAL Direct Submission
TITLE Submitted (08-JUN-2001) DOE Joint Genome Institute, 2800 Mitchell
AUTHORS Drive, Walnut Creek, CA 94598, USA
COMMENT On Jun 8, 2001 this sequence version replaced gi:13470154.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40.99.9% of Sequence;
Estimated Total Number of Errors is 0.1.
STS Content:
WT-4499 G02896
SHGC-84555 G51780
SHGC-101777 G56263.
Location/Qualifiers
1.146206
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTD-2377N4"
BASE COUNT 46774 a 28468 c 27627 g 43337 t
ORIGIN
Query Match 7.3%; Score 250.8; DB 9; Length 146206;
Best Local Similarity 92.3%; Pred. No. 6.3e-65;
Matches 264; Conservative 0; Mismatches 22; Indels 0; Gaps 0;
QY 412 GCGTGGAGGATGATACGAGGGGCTCCCACTGACACGCGGCGCCAGATGAGG 471
DB 96376 GCGTGGAGGATGATACGAGGGGCTCCCACTGACACGCGGCGCCAGATGAGG 96317
QY 472 GCTCTGTGACACCCATCACCAGGGGCTTCCACTGCTGCGGGCTTCCGGGCTGCGCC 531
DB 96316 GCTCTGTGACACCCATCACCAGGGGCTTCCACTGCTGCGGGCTTCCGGGCTGCGCC 96257
QY 532 TCGGAGACCGCTGTGAGAGGACCTATGCTACGACTGTCTCAGAGATGCCAGTGC 591
DB 96256 TCGGAGACCGCTGTGAGAGGACCTATGCTACGACTGTCTCAGAGATGCCAGTGC 96197
QY 592 CAGATNGAGCCACCTGACACGACGACGAGGGGATGCGGCTGCCACAGATGACAC 651
DB 96196 CAGATNGAGCCACCTGACACGACGACGAGGGGATGCGGCTGCCACAGATGACAC 96137
QY 652 GGAGCCTTGTGAGATCTTGTCTCCTCTGTAACATGATGTCAC 697
DB 96136 GGAGCCTTGTGAGATCTTGTCTCCTCTGTAACATGATGTCAC 96091
RESULT 14
AC008566 152765 bp DNA linear HTG 20-APR-2001
LOCUS Homo sapiens chromosome 5 clone CTC-540P9, WORKING DRAFT SEQUENCE,
DEFINITION 36 unordered pieces.
AC008566
AC008566.4 GI:13699384
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
KEYWORDS Homo sapiens.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
AUTHORS 1 (bases 1 to 152765)
JOURNAL DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 5
REFERENCE 2 (bases 1 to 152765)
AUTHORS DOE Joint Genome Institute.
JOURNAL Direct Submission
TITLE Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
AUTHORS Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT On Apr 20, 2001 this sequence version replaced gi:7708954.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov

Project Information
Center Project Name: 392909, H410
Center Clone name: CIT-HSPC_540P9

Summary Statistics
Consensus quality: 121755 bases at least Q40.
Consensus quality: 133065 bases at least Q30
Consensus quality: 136597 bases at least Q20
Estimated insert size: 120000; pulse field gel estimation
Estimated insert size: 149265; sum-of-contigs estimation
Quality coverage: 6.44 in Q20 bases; pulse field gel estimation
Quality coverage: 5.18 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 36 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
1073 1072: contig of 1072 bp in length
1173 1172: gap of unknown length
1173 2239: contig of 1067 bp in length
2240 2339: gap of unknown length
2340 3463: contig of 1124 bp in length
3464 3563: gap of unknown length
3564 4791: contig of 1228 bp in length
4792 4891: gap of unknown length
4892 5928: contig of 1037 bp in length
5929 6029: gap of unknown length
6029 7174: contig of 1146 bp in length
7175 7274: gap of unknown length
7275 8490: contig of 1216 bp in length
8491 8590: gap of unknown length
8591 9611: contig of 1020 bp in length
9611 9710: gap of unknown length
9711 10819: contig of 1109 bp in length
10820 10919: gap of unknown length
10920 12083: contig of 1164 bp in length
12084 12183: gap of unknown length
12184 13307: contig of 1124 bp in length
13308 13407: gap of unknown length
13408 14635: contig of 1228 bp in length
14636 14735: gap of unknown length
14736 15839: contig of 1104 bp in length
15840 15939: gap of unknown length
15940 16982: contig of 1043 bp in length
16983 17082: gap of unknown length
17083 18125: contig of 1043 bp in length
18126 18225: gap of unknown length
18226 19372: contig of 1147 bp in length
19373 19473: gap of unknown length
19473 20479: contig of 1007 bp in length
20480 21586: gap of unknown length
21587 21586: contig of 1007 bp in length
21587 23352: gap of unknown length
23352: contig of 1666 bp in length


```

QY 472 GCTCTGTGCAACCCCATCAGCGGGGCTTGCCACTGTGCTGCGGGCTTCGGGGGCTGCGGC 531
    |||||||
Db 94590 GCTCTGTGCAACCCCATCAGCGGGGCTTGCCACTGTGCTGCGGGGCTTCGGGGGCTGCGGC 94531
    |||||||
QY 532 TGGCAGGACCCGCTGTGAGCAGGAGCACCCTATGTTAACGACTGTCAATCAGAGATGCCAGTGC 591
    |||||||
Db 94530 TGGCAGGACCCGCTGTGAGCAGGAGCACCCTATGTTAACGACTGTCAATCAGAGATGCCAGTGC 94471
    |||||||
QY 592 CAGAAATGGAGCCACCTGCGACACGTCACGGGGGAAATGCGCTGCCACACAGATACACC 651
    |||||||
Db 94470 CAGAAATGGAGCCACCTGCGACACGTCACGGGGGAAATGCGCTGCCACACAGATACACC 94411
    |||||||
QY 652 GGAGCCTTCTGTGAGGATCTTTGTCCTCCTGTAAACATGGTCCAC 697
    |||||||
Db 94410 GGAGCCTTGTAAATCATGCTGCCCCAGCAGCAGAGAGCCAC 94365
    |||||||

```

Search completed: May 9, 2003, 10:55:18
 Job time : 10719 secs

GenCore version 5.1.4.D5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 9, 2003, 04:47:55 ; Search time 4559 Seconds

(without alignments)
12159.937 Million cell updates/sec

Title: US-10-092-390-1

Perfect score: 3423
Sequence: 1 atggtatttttttgtaactc.....gcagcagcagcagtgatga 3423

Scoring table: IDENTITY_NUC
Gapop 10.0 ; Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vit:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_frod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1334.2	39.0	3162	11	BC029999	BC029999 Homo sapi
2	456.6	13.3	755	12	BC828819	BC828819 602751395
3	433	12.6	565	14	BM719978	BM719978 UI-E-EJ0-
4	432	12.6	598	14	BM676825	BM676825 UI-E-EJ0-
5	379	11.1	538	10	BM762574	BM762574 BM762574
6	370	10.8	715	9	AI958909	AI958909 fd05g03.y

RESULT 1	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	REMARK	COMMENT
BC029999	BC029999	Homo sapiens, clone IMAGE:4156083, mRNA.	BC029999	BC029999.1	GI:20455873	Homo sapiens.	1	(bases 1 to 3162)	Submitted (06-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA			
BC029999	BC029999	Homo sapiens, clone IMAGE:4156083, mRNA.	BC029999	BC029999.1	GI:20455873	Homo sapiens.	2	(bases 1 to 3162)	Submitted (06-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA			
BC029999	BC029999	Homo sapiens, clone IMAGE:4156083, mRNA.	BC029999	BC029999.1	GI:20455873	Homo sapiens.	3	(bases 1 to 3162)	Submitted (06-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA			
BC029999	BC029999	Homo sapiens, clone IMAGE:4156083, mRNA.	BC029999	BC029999.1	GI:20455873	Homo sapiens.	4	(bases 1 to 3162)	Submitted (06-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA			
BC029999	BC029999	Homo sapiens, clone IMAGE:4156083, mRNA.	BC029999	BC029999.1	GI:20455873	Homo sapiens.	5	(bases 1 to 3162)	Submitted (06-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA			
BC029999	BC029999	Homo sapiens, clone IMAGE:4156083, mRNA.	BC029999	BC029999.1	GI:20455873	Homo sapiens.	6	(bases 1 to 3162)	Submitted (06-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA			
BC029999	BC029999	Homo sapiens, clone IMAGE:4156083, mRNA.	BC029999	BC029999.1	GI:20455873	Homo sapiens.	7	(bases 1 to 3162)	Submitted (06-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA			
BC029999	BC029999	Homo sapiens, clone IMAGE:4156083, mRNA.	BC029999	BC029999.1	GI:20455873	Homo sapiens.	8	(bases 1 to 3162)	Submitted (06-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA			
BC029999	BC029999	Homo sapiens, clone IMAGE:4156083, mRNA.	BC029999	BC029999.1	GI:20455873	Homo sapiens.	9	(bases 1 to 3162)	Submitted (06-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA			
BC029999	BC029999	Homo sapiens, clone IMAGE:4156083, mRNA.	BC029999	BC029999.1	GI:20455873	Homo sapiens.	10	(bases 1 to 3162)	Submitted (06-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA			
BC029999	BC029999	Homo sapiens, clone IMAGE:4156083, mRNA.	BC029999	BC029999.1	GI:20455873	Homo sapiens.	11	(bases 1 to 3162)	Submitted (06-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA			
BC029999	BC029999	Homo sapiens, clone IMAGE:4156083, mRNA.	BC029999	BC029999.1	GI:20455873	Homo sapiens.	12	(bases 1 to 3162)	Submitted (06-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA			
BC029999	BC029999	Homo sapiens, clone IMAGE:4156083, mRNA.	BC029999	BC029999.1	GI:20455873	Homo sapiens.	13	(bases 1 to 3162)	Submitted (06-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA			
BC029999	BC029999	Homo sapiens, clone IMAGE:4156083, mRNA.	BC029999	BC029999.1	GI:20455873	Homo sapiens.	14	(bases 1 to 3162)	Submitted (06-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA			
BC029999	BC029999	Homo sapiens, clone IMAGE:4156083, mRNA.	BC029999	BC029999.1	GI:20455873	Homo sapiens.	15	(bases 1 to 3162)	Submitted (06-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA			
BC029999	BC029999	Homo sapiens, clone IMAGE:4156083, mRNA.	BC029999	BC029999.1	GI:20455873	Homo sapiens.	16	(bases 1 to 3162)	Submitted (06-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA			
BC029999	BC029999	Homo sapiens, clone IMAGE:4156083, mRNA.	BC029999	BC029999.1	GI:20455873	Homo sapiens.	17	(bases 1 to 3162)	Submitted (06-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA			
BC029999	BC029999	Homo sapiens, clone IMAGE:4156083, mRNA.	BC029999	BC029999.1	GI:20455873	Homo sapiens.	18	(bases 1 to 3162)	Submitted (06-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA			
BC029999	BC029999	Homo sapiens, clone IMAGE:4156083, mRNA.	BC029999	BC029999.1	GI:20455873	Homo sapiens.	19	(bases 1 to 3162)	Submitted (06-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA			
BC029999	BC029999	Homo sapiens, clone IMAGE:4156083, mRNA.	BC029999	BC029999.1	GI:20455873	Homo sapiens.	20	(bases 1 to 3162)	Submitted (06-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA			
BC029999	BC029999	Homo sapiens, clone IMAGE:4156083, mRNA.	BC029999	BC029999.1	GI:20455873	Homo sapiens.	21	(bases 1 to 3162)	Submitted (06-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA			
BC029999	BC029999	Homo sapiens, clone IMAGE:4156083, mRNA.	BC029999	BC029999.1	GI:20455873	Homo sapiens.	22	(bases 1 to 3162)	Submitted (06-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA			
BC029999	BC029999	Homo sapiens, clone IMAGE:4156083, mRNA.	BC029999	BC029999.1	GI:20455873	Homo sapiens.	23	(bases 1 to 3162)	Submitted (06-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA			
BC029999	BC029999	Homo sapiens, clone IMAGE:4156083, mRNA.	BC029999	BC029999.1	GI:20455873	Homo sapiens.	24	(bases 1 to 3162)	Submitted (06-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA			
BC029999	BC029999	Homo sapiens, clone IMAGE:4156083, mRNA.	BC029999	BC029999.1	GI:20455873	Homo sapiens.	25	(bases 1 to 3162)	Submitted (06-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA			
BC029999	BC029999	Homo sapiens, clone IMAGE:4156083, mRNA.	BC029999	BC029999.1	GI:20455873	Homo sapiens.	26	(bases 1 to 3162)	Submitted (06-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA			
BC029999	BC029999	Homo sapiens, clone IMAGE:4156083, mRNA.	BC029999	BC029999.1	GI:20455873	Homo sapiens.	27	(bases 1 to 3162)	Submitted (06-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA			
BC029999	BC029999	Homo sapiens, clone IMAGE:4156083, mRNA.	BC029999	BC029999.1	GI:20455873	Homo sapiens.	28	(bases 1 to 3162)	Submitted (06-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA			
BC029999	BC029999	Homo sapiens, clone IMAGE:4156083, mRNA.	BC029999	BC029999.1	GI:20455873	Homo sapiens.	29	(bases 1 to 3162)	Submitted (06-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA			
BC029999	BC029999	Homo sapiens, clone IMAGE:4156083, mRNA.	BC029999	BC029999.1	GI:20455873	Homo sapiens.	30	(bases 1 to 3162)	Submitted (06-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA			
BC029999	BC029999	Homo sapiens, clone IMAGE:4156083, mRNA.	BC029999	BC029999.1	GI:20455873	Homo sapiens.	31	(bases 1 to 3162)	Submitted (06-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA			
BC029999	BC029999	Homo sapiens, clone IMAGE:4156083, mRNA.	BC029999	BC029999.1	GI:20455873	Homo sapiens.	32	(bases 1 to 3162)	Submitted (06-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA			
BC029999	BC029999	Homo sapiens, clone IMAGE:4156083, mRNA.	BC029999	BC029999.1	GI:20455873	Homo sapiens.	33	(bases 1 to 3162)	Submitted (06-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA			
BC029999	BC029999	Homo sapiens, clone IMAGE:4156083, mRNA.	BC029999	BC029999.1	GI:20455873	Homo sapiens.	34	(bases 1 to 3162)	Submitted (06-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA			
BC029999	BC029999	Homo sapiens, clone IMAGE:4156083, mRNA.	BC029999	BC029999.1	GI:20455873	Homo sapiens.	35	(bases 1 to 3162)	Submitted (06-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA			
BC029999	BC029999	Homo sapiens, clone IMAGE:4156083, mRNA.	BC029999	BC029999.1	GI:20455873	Homo sapiens.	36	(bases 1 to 3162)	Submitted (06-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA			
BC029999	BC029999	Homo sapiens, clone IMAGE:4156083, mRNA.	BC029999	BC029999.1	GI:20455873	Homo sapiens.	37	(bases 1 to 3162)	Submitted (06-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA			
BC029999	BC029999	Homo sapiens, clone IMAGE:4156083, mRNA.	BC029999	BC029999.1	GI:20455873	Homo sapiens.	38	(bases 1 to 3162)	Submitted (06-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA			
BC029999	BC029999	Homo sapiens, clone IMAGE:4156083, mRNA.	BC029999	BC029999.1	GI:20455873	Homo sapiens.	39	(bases 1 to 3162)	Submitted (06-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA			
BC029999	BC029999	Homo sapiens, clone IMAGE:4156083, mRNA.	BC029999	BC029999.1	GI:20455873	Homo sapiens.	40	(bases 1 to 3162)	Submitted (06-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA			
BC029999	BC029999	Homo sapiens, clone IMAGE:4156083, mRNA.	BC029999	BC029999.1	GI:20455873	Homo sapiens.	41	(bases 1 to 3162)	Submitted (06-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA			
BC029999	BC029999	Homo sapiens, clone IMAGE:4156083, mRNA.	BC029999	BC029999.1	GI:20455873	Homo sapiens.	42	(bases 1 to 3162)	Submitted (06-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA			
BC029999	BC029999	Homo sapiens, clone IMAGE:4156083, mRNA.	BC029999	BC029999.1	GI:20455873	Homo sapiens.	43	(bases 1 to 3162)	Submitted (06-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA			
BC029999	BC029999	Homo sapiens, clone IMAGE:4156083, mRNA.	BC029999	BC029999.1	GI:20455873	Homo sapiens.	44	(bases 1 to 3162)	Submitted (06-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA			
BC029999	BC029999	Homo sapiens, clone IMAGE:4156083, mRNA.	BC029999	BC029999.1	GI:20455873	Homo sapiens.	45	(bases 1 to 3162)	Submitted (06-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA			

ALIGNMENTS

RESULT 1
LOCUS BC029999 3162 bp mRNA
DEFINITION Homo sapiens, clone IMAGE:4156083, mRNA.
ACCESSION BC029999
VERSION BC029999.1 GI:20455873
KEYWORDS
SOURCE
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
1 (bases 1 to 3162)
Strausberg, R.
Direct Submission
Submitted (06-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK
NHL-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: David N. Louis, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: amg@bcm.tmc.edu

Db 748 GCCACCC 754

RESULT 3
LOCUSBM719978 565 bp mRNA linear EST 01-MAR-2002
UI-E-EJ0-ahu-1-16-0-UI.r1 UI-E-EJ0 Homo sapiens cDNA cloneDEFINITION
BM719978 UI-E-EJ0-ahu-1-16-0-UI 5', mRNA sequence.ACCESSION
BM719978 GI:19038910VERSION
BM719978.1KEYWORDS
human.ORGANISM
Homo sapiensEukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.REFERENCE
1 (bases 1 to 565)
Bonaldo, M.F., Lennon, G. and Soares, M.B.AUTHORS
Normalization and subtraction: two approaches to facilitate gene
discoveryTITLE
Genome Res. 6 (9), 791-806 (1996)JOURNAL
MEDLINE
97044477COMMENT
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.eduTissue Procurement: Dr. Gregg Hageman
CDNA Library preparation: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.reagen.com).

Seq primer: M13 Reverse.

Location/Qualifiers

1..565

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="UI-E-EJ0-ahu-1-16-0-UI"

/clone_lib="UI-E-EJ0"

/tissue_type="fetal eyes, lens, eye anterior segment,
optic nerve, retina, Retina Foveal and Macular, RPE and
Choroid"

/dev_stage="fetal and adult"

/lab_host="DH10B (Life Technologies) (T1 phage resistant)"

/note="Organ: eye; Vector: pRT73-Pac (Pharmacia) with a
modified polylinker; Site.1: EcoR I; Site.2: Not I;
UI-E-EJ0 is a subtracted cDNA library constructed
according to Bonaldo, Lennon and Soares, Genome Research,
6:791-806, 1996. First strand cDNA synthesis was primed
with an oligo-dT primer containing a Not I site. Double
stranded cDNA was ligated to an EcoR I adaptor, digested
with Not I, and cloned directionally into pRT73-Pac
vector. The oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tags for this library are: fetal eyes, ACATACAGCA
; lens, GCATACAGCA; eye anterior segment, AATGCCGAT;
optic nerve, CCATACAGCA; retina, CCGCG; Retina Foveal and
Macular, GTCC; RPE and Choroid, ACCCTA. This library was
created for the program, Gene Discovery in the Visual
System, supported by National Eye Institute (NEI)."

BASE COUNT 179 a 122 c 139 g 125 t

ORIGIN

Query Match 12.6%; Score 433; DB 14; Length 565;

Best Local Similarity 100.0%; Pred. No. 9.5e-116;

Matches 433; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2800 ACCCAGTGTGCACATCCCTCAGTCACACAGGAGAGATGATCAGCAAGTCA 2859

|||||

Db 1 ACCCAGTGTGCACATCCCTCAGTCACACAGGAGAGATGATCAGCAAGTCA 60

|||||

|||||

QY 2860 AAAAAACATCAATCTTTTGTGATCTTAAAAATGTGAACCTGGGAAGAGAGCCCTGTG 2919

Db 61 AAAAAACATCAATCTTTTGTGATCTTAAAAATGTGAACCTGGGAAGAGAGCCCTGTG 120

QY 2920 GGGGACGTCAGTGGGACATGTCGGGCTGACTGGGAAACATGGCGGCTACCTCAGACACTC 2979

Db 121 GGGGACGTCAGTGGGACATGTCGGGCTGACTGGGAAACATGGCGGCTACCTCAGACACTC 180

QY 2980 GGTCTTTTGTGACTTGAACAGAGCTATATGGAATCTTAAAAAGACCTGGGAAAGAT 3039

Db 181 GGTCTTTTGTGACTTGAACAGAGCTATATGGAATCTTAAAAAGACCTGGGAAAGAT 240

QY 3040 TCTGAATATATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 3099

Db 241 TCTGAATATATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 300

QY 3100 AAGACCCACCTGATCTTATCCCAAGGAAAGCTGAGAGTGTGATGAGATGAATATG 3159

Db 301 AAGACCCACCTGATCTTATCCCAAGGAAAGCTGAGAGTGTGATGAGATGAATATG 360

QY 3160 CCGGACAGAGAGATTCCTCATATGACAGATCAATCAATCAATCAATCAATCAATCAAT 3219

Db 361 CCGGACAGAGAGATTCCTCATATGACAGATCAATCAATCAATCAATCAATCAATCAAT 420

QY 3220 GTCATGAGATTTG 3232

Db 421 GTCATGAGATTTG 433

RESULT 4

BM676825/c 598 bp mRNA linear EST 27-FEB-2002

UI-E-EJ0-ahu-1-16-0-UI.s2 UI-E-EJ0 Homo sapiens cDNA clone

DEFINITION
UI-E-EJ0-ahu-1-16-0-UI 3', mRNA sequence.ACCESSION
BM676825VERSION
BM676825.1 GI:18986721KEYWORDS
EST.SOURCE
human.ORGANISM
Homo sapiensEukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.REFERENCE
1 (bases 1 to 598)
Bonaldo, M.F., Lennon, G. and Soares, M.B.AUTHORS
Normalization and subtraction: two approaches to facilitate gene
discoveryTITLE
Genome Res. 6 (9), 791-806 (1996)JOURNAL
MEDLINE
97044477COMMENT
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.eduTissue Procurement: Dr. Gregg Hageman
CDNA Library preparation: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.reagen.com).

Seq primer: M13 forward

POLY-A-yes.

Location/Qualifiers

1..598

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="UI-E-EJ0-ahu-1-16-0-UI"

/clone_lib="UI-E-EJ0"

/tissue_type="fetal eyes, lens, eye anterior segment,
optic nerve, retina, Retina Foveal and Macular, RPE and
Choroid"

/dev_stage="fetal and adult"

Db	120	CTTTACCAATCCCAATTATACACACTTAGCCAGTGTGCACACATCCCTCANTGTAAACA	179
Oy	2832	CAGGACACAGGATGACTGTGCACGAAAGTCAAAAAACATACACTGTTTGTGAATCTTAAAAA	2891
Db	180	CAGGACACAGGATGACCAATTCGAAAGTCAAAAAAACATACAGCTGTTTGTGAATCTTAAAAA	239
Oy	2892	TGTGAACCTGGGAAGAGAGGCGCTGTGGGGAGCTGCA-CTGGGACATTTGGCGGCTGACT	2950
Db	240	TGTGAATCCAGGAAAGAGAGGACACTTGTGTGACTGTGCACTTGGCAATTTGCCAGCTGACT	299
Oy	2951	GGAAACATGGCGGCTACCTCAACGAGCTCGGTGCTTTTGGACTTACACAGAAGCTATGAG	3010
Db	300	GGAAACAGAAGAGCGTACCTCAATGAGCTTGGGTGCTTGGGGGTGGACAGAAGACTACATAGG	359
Oy	3011	GAAAACTCTTAAAGACACTCGGAAAGAA-TTCTGAATATTAATTCAAGTAACTGCTCCCTTA	3069
Db	360	GAAATCCTTAAAGATGTGGGGAAGAGCTGTGAATTAATTCAAGCACTTGCTCCCTTA	419
Oy	3070	AGCAATTTCTGAGAACCCATATGCCACTATTTAAAGAACCACTGTACTTATCCCGAAAAAGC	3129
Db	420	AGCACTCTTGAAAAACCCATATGCCCACTTAAAGAACCGCGCTGCACCTCTGCTTAAAGC	479
Oy	3130	TCAGAGTGTGGTTA-TGTGAGATGAATAATCGCGGACAGAAAGATTTCCCATATATGACAG	3187
Db	480	TCCGAGTGGCGGTACACTGTGGAGATATAAGTCCGCGCGCGGAGAGACTCCCCATATATGACAG	538

[illegible]

QY	3152	GGCAGCAAGAGATTTCCCATTTGCGAGATCAATTAATCACTTACGACCAAGAAATGT	3221
Db	301	GGCTGGAGGGAGCTCCCATACGCGAGATCAATTAATCACTTACGCAACAGAAATGT	360
QY	3222	CTATGAATTTGAACTTACAGAGTGAAGTGTTCACAGAGATTAATTAATGGGCGCTCT	3281
Db	361	CTATGAATTTGAACTTACAGAGTGAAGTGTTCACAGAGATTAATTAATGGGCGCACT	420
QY	3282	CTCCAGAGATTCATATGA	3299
Db	421	CACATGAGATTCCTATGA	438
RESULT 8			
BF529240			
LOCUS			
DEFINITION	BF529240	779 bp	mRNA linear EST 11-Dec-2000
ACCESSION	602041695F1	NCI_CGAP_Brn67	Homo sapiens cDNA clone IMAGE:4179646
VERSION	5', mRNA sequence.		
KEYWORDS	BF529240		
SOURCE	BF529240.1	GI:11616603	
ORGANISM	Est.		
REFERENCE	human.		
AUTHORS	Homo sapiens		
TITLE	Eukaryota: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
JOURNAL	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
COMMENT	NIH-MGC http://mgc.nci.nih.gov/.		
	National Institutes of Health, Mammalian Gene Collection (MGC)		
	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: cgapbs-remail.nih.gov		
	Tissue Procurement: David N. Louis, M.D.		
	cDNA Library Preparation: Life Technologies, Inc.		
	CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)		
	DNA sequencing by: Incyte Genomics, Inc.		
	Clone distribution: MGC clone distribution information can be		
	found through the I.M.A.G.E. Consortium/LNL at:		
	http://image.lnl.gov		
	Plate: LHAM9489	row: m	column: 23
	High quality sequence stop: 737.		
FEATURES			
source			
	1..779		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:4179646"		
	/clone_id="NCI_CGAP_Brn67"		
	/tissue_type="anaplastic oligodendroglioma with 1p/19q		
	loss"		
	/lab_host="DH10B (T1 phage-resistant)"		
	/note="Organ: brain; Vector: PCMV-SpOrf6; Site_1: NotI;		
	Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.		
	Average insert size 2.3 kb. Constructed by Life		
	Technologies. Note: this is a NCI_CGAP library."		
BASE COUNT	132 a	245 c	239 g
ORIGIN			163 t
Query Match	10.7%;	Score 366.8;	DB 12; Length 779;
Best Local Similarity	70.2%;	Pred. No. 3.8e-96;	
Matches 550;	Conservative 0;	Mismatches 227;	Indels 7; Gaps 4;
QY	1159	TGCAAGCCGGGCTGTGACGACTCTACTCTGTATATGACATGTTCCTCTGATTTACGGG	1218
Db	1	TGCCAGCCAGGCTGTGCTGTGTCACACACGCAATGATCTCCCTGGTGGCTACTATTCG	60
QY	1219	GAAAGTTCGACAGATGTGAGAGTGGCCAAATATGGAGGAGACAGTGGACATGTCAGTGA	1278
Db	61	GATGGCTCCAGCTGCTCTTCGACCTGTATAGATATGGCGCCGACTGCCACACATCATCTGG	120
QY	1279	AAGTGCACCTGTGCCCCAGATTCGCAAGAAATGATGACTCTTACCCCATGCCCCTCTGGGA	1338
Db	121	GGCTGCACCTGTGCTCGCGGCTTATATGGAGAGGCTGTGCGGTTCTGTGACAGACAGG	180
QY	1339	ACCTATGGGATTAATCTTCTCTGCTGCTGTGGCTGTAAAAATGATGACAGTGTGCTTCT	1398

Db	181	ACGATGCGCCCACTGCTCGCATTTGTAGCTGTAACTAATGATGGGACACCTGCTCCCA	240
Qy	1399	GTGAGGGGCTTGTGATTGCAAGGCAAGGCTGACGCGGGGTGGAGTCTCCATCAGATGT	1458
Db	241	GTAGATGGCTCTGTACTGTGCAAGGAAAGGGGCGAGGGCTGTGAGTGCACCTGTGCATGT	300
Qy	1459	CCGAGTGGCACATGGGGGCTTTGGCTGTAACTTAACTACCATCCAGTGGCTTCACAGGGAGGAC	1518
Db	301	CCGAGTGGGAGCTGGGGGCTGAATGTGCAACAGAGACTCTACCTGTGGCCAAATGGGACACC	360
Qy	1519	TGCAACACCTTGAGAGGAGCACTGCACGTGTGCACCTGATGGCGCGGAGAAATGCA	1578
Db	361	TGCAGCCCCCAATAGAGGGCTCTCTCTCTGCACTCTGTGCTGGGAGACACCTGTAG	420
Qy	1579	CTTCCCTCCGAGGATGGGACGAGGGGCTGAACTGTGTGAGCGGTGGGACGTGCAGCCAC	1638
Db	421	CTGCTTTCCTCCGATGGACATTTGGCTGAACTGCACTGAACTGTGACCTGTGACGACAT	480
Qy	1639	GCAGATGGCTGCACACCTTACACAGGCGCCATTTGCCGCTCTCCCGGAGTGTCAAGTGT	1698
Db	481	GCTGATGGATGTGAGCCCGTACAGAGGCCACGTCTCTCTGTGGCCGAGTGAACAGGACAT	540
Qy	1699	CACGTGTGACACGCTGTGTGTGTGAGGAGAGCTGGGGGCCCCCACTGCTCCCTGCTGTAC	1758
Db	541	CGCTGTGACACGACCTGT	600
Qy	1759	TGTAAATATGGGGGCTTACATGCTCCCTCTATGATGATGATGATGATGATGATGATGATG	1818
Db	601	TGTGATGATGAGGGCTCTGT	660
Qy	1819	CGAGGACACCTTGT	1878
Db	661	CGAGG-ACCTTATGTGCAAGATGTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	716
Qy	1879	ACATGCCCAAGTGGCTTACAGAGGCGGCGGCTTCACACATACAGCGGCGCTGTGTGAC	1938
Db	717	CCATGCC-CTCTGT	773
Qy	1939	TGCT 1942	
Db	774	TGCT 777	

Email: genome-res@gsc.riken.go.jp,
URL: <http://genome.gsc.riken.go.jp/>,
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh
, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi, K., Fujisake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,
Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuzura
, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and
Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara
, Y. and Hayashizaki, Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Yamanaka, I., Kiyosawa, H., Kondo, S., Saito, T., Shinagawa, A., Aizawa
, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K., Arai, T.,
Ishii, Y. and Hayashizaki, Y.
Mapping of 19032 mouse cDNAs on mouse chromosomes. J. Struct.
Funct. Genomics 2 pre, L72-L85 (2001)
Please visit our web site (<http://genome.gsc.riken.go.jp>) for
further details.
e mouse tissues.

FEATURES	source	Location/Qualifiers
		1. .641
		/organism="Mus musculus"
		/strain="C57BL/6J"
		/db_xref="taxon:10090"
		/clone="6530404N3"
		/clone_id="RIKEN full-length enriched, 10 days neonate cerebellum"
		/sex="mixed"
		/tissue_type="cerebellum"
		/dev_stage="10 days neonate"
		/lab_host="DH10B"
		/note="Site_1: SalI. Site_2: BamHI. cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer 15' GAGGAGAGAGGATCCACAGCCTCTTTTCTTTTCTTTTCTTTVN 3', cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adaptor of sequence 15' GAGGAGAGAGATTCGAGATTAATTAATCCCCCCCCCC 3'. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pluescript KS(+) after bulk excision from Lambda FliC I. Cloning sites, 5' end: SalI; 3' end: BamHI.
BASE COUNT	113 a 204 c 194 g 130 t	
ORIGIN		
Query Match	10.5%: Score 359.6; DB 10; Length 641;	
Best Local Similarity	72.7%: Pred. No. 4.4e-94;	
Matches 464, Conservative 0; Mismatches 174; Indels 0; Gaps 0;		
QY 94	AATGTGTATACCACTGGGAAGCTACTACTGACTGTGCAAGAGCTCATACCCACATCCC 153	
Db 1	AATGTGTATACCACTGGGAAGAGCTATCCCGTAGCTGTGCATGAGTCTTATGCAACACCCC 60	
QY 154	TTTGTATCAAAATTTACTACAGCAGTCGACACTCATCTTAAACTGGTTTAAATGCAAGCGGG 213	
Db 61	TTTGTATCAATTTACTATACACAGATGTGCAGACATCCCTCAACTGGTTTAAAGTATACCGGG 120	
QY 214	CACAGAGTCACACTATCGACAGAGCCATATGACATGGGGGAAGAGACATCTTATAGCGGCAG 273	

```

Db 121 CACCGGATGAGTATAGACCGGCTATAGCGGGCCCGACCATGTACCGCGGAG 180
Oy 274 TGTCACTGTTCCTCGATTTATGAACGGGGAATGTGTCTCCCACTGTCTGAT 333
Db 181 TCCCAATGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240
Oy 334 AATGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 393
Db 241 GAGTGTATGACGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300
Oy 394 GGGACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 453
Db 301 GGGCTGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 360
Oy 454 TGGCAATGCAAAATGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 513
Db 361 TGTCACTGTGACAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420
Oy 514 GGGTCCGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 573
Db 421 GGGTCCGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 480
Oy 574 CATCAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 633
Db 481 CAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 540
Oy 634 TGGCCACCAAGATACACCGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 693
Db 541 TGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600
Oy 694 CCACAGTGTGACAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 731
Db 601 GCTCACTGTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 638

```

```

RESULT 10
LOCUS H15471 488 bp mRNA linear EST 27-JUN-1995
DEFINITION ym39b02.r1 Soares infant brain INIB Homo sapiens cDNA clone
IMAGE:49332 5, mRNA sequence.
ACCESSION H15471
VERSION H15471.1 GI:880291
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 488)
AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman,
M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J.,
Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevisan, E., Waterston,
R., Williamson, A., Wohldmann, P. and Wilson, R.
The WashU-Merck EST Project
Unpublished (1995)
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 2474
High quality sequence stops: 305
Source: IMAGE Consortium, LNL
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 2474 Std Error: 0.00
Seq primer: M13RP1
High quality sequence stop: 305.
Location/Qualifiers
1.488
/organism="Homo sapiens"
/db_xref="GDB:421873"

```

```

/db_xref="taxon:9606"
/clone="IMAGE:49332"
/clone_lib="Soares infant brain INIB"
/sex="female"
/dev_stage="73 days post natal"
/lab_host="DH10B (ampicillin resistant)"
/organ="Whole brain; Vector: Latmid BA; Site:1: Not
I; Site:2: Hind III; 1st strand cDNA was primed with a Not
I - oligo(dT) primer [5',
AACTGGAAGAATGCGCGCGGAGGAGATTTTCTTTTCTTTTCTTTTCTTTTCTTTT
double-stranded cDNA was ligated to Hind III adaptors
(Pharmacia), digested with Not I and directionally cloned
into the Not I and Hind III sites of the Latmid BA vector.
Library went through one round of normalization. Library
constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 144 a 117 c 112 g 113 t 2 others
ORIGIN
Query Match 10.4%; Score 354.6; DB 14; Length 488;
Best Local Similarity 93.8%; Pred. No. 1,1e-92;
Matches 425; Conservative 0; Mismatches 19; Indels 9; Gaps 5;

```

```

Oy 2628 CATTAATTTATAGACACAGAGGAAAGAAATCAACATGCCAGCTTACTTAC 2687
Db 3 CATTAATTTATAGACACAGAGGAAAGAAATCAACATGCCAGCTTACTTAC 62
Oy 2688 CCTGCTATGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2747
Db 63 CCTGCTATGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 122
Oy 2748 TGGTGAAGAGCTAATAGGCACTTCTTACCAATCCAGTACACAGCTGCTG 2807
Db 123 TGGTGAAGAGCTAATAGGCACTTCTTACCAATCCAGTACACAGCTGCTG 182
Oy 2808 TGCACATCCCTCAGCTGACAGAGGAGAGATGCTGCTGCTGCTGCTGCTG 2867
Db 183 TGCACATCCCTCAGCTGACAGAGGAGAGATGCTGCTGCTGCTGCTGCTG 242
Oy 2868 TCAACTGTTTGAATCTTAAATGTGAACCTTGGGAGAGAGGCTGCTGCTG 2926
Db 243 TCAACTGTTTGAATCTTAAATGTGAACCTTGGGAGAGAGGCTGCTGCTG 302
Oy 2927 GCACGTGGAGATTCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2984
Db 303 GCACGTGGAGATTCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 362
Oy 2985 TTTTGACTTGAACAGAGCTATAT-GGAAAAATCTTAAAGAGCT--TGGAAAGAA 3040
Db 363 TTTTGACTTGAACAGAGCTATATTTGGGAAAAATCTTAAAGAGAGCTGGGAAAGATT 422
Oy 3041 CTGAATATATAT--CAAGTACTGCTCCCTAAG 3071
Db 423 TTGAATATATATTTCAAGTTAAGTCTTCTCTAAG 455

```

```

RESULT 11
LOCUS BF686873 584 bp mRNA linear EST 22-DEC-2000
DEFINITION 602102830P1 NCI_CGAP_Kid14 Mus musculus cDNA clone IMAGE:4221079
IMAGE:4221079
ACCESSION BF686873
VERSION BF686873
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 584)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE NIH-MGC
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov

```


Db 414 CAGTACTGAGAGCTTGTACCTGCGACGAGGCTGCTGCTGTCACCACTGCAATGAACTCCT 473
 Oy 1199 GTTCTCCGATGATCTACAGGGGAGAGCTGCGACAGATGCGACGCTGCCAAATGGGGCG 1258
 Db 474 GCCCTGTGGCTACTATGCGATGGCTGCGACGCTGCTTGCACCTGTGCGAATGCGCGCG 533
 Oy 1259 ACTGTGACAGTGTGACTGGAAGTGCACCTGTGCGCCAGAGATTTAAAGGAATGACTGCT 1318
 Db 534 ACTGCGACAGCATCACTGCGGGGCTGCGACTGTGCTCCGGCTTCAATGGAGAGCTGTG 593
 Oy 1319 CTACCCCATGCCCCCTGTGGGAACCTATGGGATACTGTTCTCTGCTGCTGCTGTAAAA 1378
 Db 594 CCGTTTCCCTGTGCGACAGGAGCTATGGCCCACTGCTGCTCAATCTTACTGTAAACA 653
 Oy 1379 ATGATGAGTGTGCTGCTCTCTGCGAGGCT-CTTGTACTTGAAGGCA--GGCTGGCAG 1435
 Db 654 ATGTGGGACCTGCTGCCAGTAGATGCTCTCTTACTGCTGAAGAAAGGGTGGCANG 713
 Oy 1436 GGGTGAAGTGTCTCATAGATGCTCCAGTGCACATGGGCTTGGCTGTACTTAACAT 1495
 Db 714 CCTTGAGTGCACCTGCGCATGTCCAGTGGGAGCTGGGGCTGAACTGCACAGAGCT 773
 Oy 1496 GCCAGTGCCTCAAGGGGAGCTGCAACACCTGAGCGGAGCTGCACGTGTGACCTG 1555
 Db 774 GCACCTGTGCCAATGGGCGAAGCTGCACCCCAATACGCGCTCCGGCTCCGAGCTCCGG 833
 Oy 1556 GATGCGCGGGGAGAAATGCGAATCTCCCTG-CCAGAGTGCAGCGTGAAGGCTGAAC 1612
 Db 834 GCTGTGTGAAACCCCTGTGAAATGCTCCCTGCCCGAAGGCGACCTTTGGGCTGAAC 891

RESULT 13
 AM58138 559 bp mRNA linear EST 25-APR-2001
 LOCUS AM58138
 DEFINITION 93921 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.
 ACCESSION AM58138
 VERSION AM58138.1 GI:7423964
 KEYWORDS EST.

SOURCE

ORGANISM Bos taurus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 Bovidae; Bovinae; Bos.

REFERENCE

AUTHORS Smith,T.P.L., Grose,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
 Casas,E., Wray,J.E., White,J., Cho,J., Fahnenkrug,S.C., Bennett,
 G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
 Pettea,G., Holt,I., Karameycheva,S., Liang,F., Quackenbush,J. and
 Keefe,J.W.

TITLE

Sequence evaluation of four pooled-tissue normalized bovine cDNA
 libraries and construction of a gene index for cattle

JOURNAL Genome Res. 11 (4), 626-630 (2001)
 MEDLINE 21180013

COMMENT

Contact: Smith TPL
 USDA, ARS, US Meat Animal Research Center
 PO Box 166, Clay Center, NE 68933-0166, USA
 Tel: 402 762 4366
 Fax: 402 762 4390
 Email: smith@email.marc.usda.gov
 Single pass sequencing. Bases called and alt-trimmed with phred
 v0.980904.e. Vector identified by cross-match with the -minscore 18
 and -minmatch 12 options.

PCR Primers

FORWARD: AGGAACAGCTATGACCAT
 BACKWARD: GTTTCCTCAGTCACGACG
 Plate: 60 row: 1 column: 12
 Seq primer: ATTAGTGCACATATATG.

FEATURES

Source

1..559
 /organism="Bos taurus"
 /db_xref="taxon:9913"
 /clone_lib="MARC 1BOV"
 /tissue_type="pooled"

/lab_host="DH10B"
 /note="Vector: pCMV SPORT6; Site.1: NotI; Site.2: SalI;
 Library made from pooled tissue from lymph node, ovary,
 fat, hypothalamus, and pituitary."

BASE COUNT 86 a 182 c 181 g 110 t

ORIGIN

Query Match 8.8%; Score 302; DB 10; Length 559;
 Best Local Similarity 71.3%; Pred. No. 3.9e-77;
 Matches 398; Conservative 0; Mismatches 160; Indels 0; Gaps 0;

Oy 1226 GCCACAGATCTGACCTGCGCCAAATGGGGCAGAGTGTGACGTGAGTGAAGATGCA 1285
 Db 2 GCCAGCTGCTGTGACACTGTCTGCAAGACGGTCCGAGATGCCACAGCATCACTGGGAGCTGA 61
 Oy 1286 CCTGTGCCCGCAGATTCAGGAATTTACTGCTCTTACCCCATGCCCCCTGTGGACCTATG 1345
 Db 62 CTTGTGCCCGCAGAGCTTCATGGGGAGGCTGTGCGACGCTCTGCGGTGGGAGCTATG 121
 Oy 1346 GGATTAAGTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1405
 Db 122 GCTCCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 181
 Oy 1406 GCTCTTGTACTTGTCAAGGAGGCTGCGACGGGGTGGAGTGTCTCATAGATGTCCAGTG 1465
 Db 182 GCTCTGACACTGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 241
 Oy 1466 GCACATGGGGCTTTGGCTGTGTAATTAATCAATGCAATGCAATGCAATGCAATGCAATGCA 1525
 Db 242 GGACGTGGGGCTGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATG 301
 Oy 1526 CCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1585
 Db 302 CCACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 361
 Oy 1586 GCCAGATGGACAGTACGGGCTGAATGTGTGACGCGTGCAGTGCAGACGACGAGATG 1645
 Db 362 GCCCGATGGACAGTACGGGCTGAATGTGTGACGCGTGCAGTGCAGACGACGAGATG 421
 Oy 1646 GCTGCCACCTTACACAGGGGCTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1705
 Db 422 GCTTGAGACCCCTGACAGCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 481
 Oy 1706 ACAGCGTGTGCTGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1765
 Db 482 ACAGCAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 541
 Oy 1766 ATGGGGCTTCATGCTGCC 1783
 Db 542 ATGGGGCTTCCTGCTGCC 559

RESULT 14

BF180097 937 bp mRNA linear EST 31-OCT-2000
 LOCUS BF180097
 DEFINITION 60180645BFL NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:4037318 5',
 mRNA sequence.

ACCESSION BF180097
 VERSION BF180097.1 GI:11058239

KEYWORDS

SOURCE

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 Mus musculus.
 house mouse.
 EST.
 BF180097
 CONTACT: Robert Strausberg, Ph.D.
 Email: cga@bbs.fda.gov
 Tissue Procurement: Lohar Hennighausen Ph.D., Robin Humphreys
 cDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

COMMENT

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 9, 2003, 07:45:02 ; Search time 403 Seconds
(without alignments)
10555.283 Million cell updates/sec

Title: US-10-092-390-1
Perfect score: 3423
Sequence: 1 atggtattcttctgaactc.....gcacagcagcagtgatgta 3423

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 783854 seqs, 621352466 residues

otal number of hits satisfying chosen parameters: 1567708

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_MA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCr_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCrUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3423	100.0	3423	9	US-10-092-390-1
2	1760	51.4	1761	9	US-10-092-390-3
3	1425.8	41.7	1448	12	US-10-105-929-9
4	608.4	17.8	3567	9	US-09-796-753-99
5	608.4	17.8	3567	9	US-09-796-753-123
6	528.6	15.4	5036	9	US-09-796-753-113
7	524.6	15.3	5197	10	US-09-833-381-1910
8	424.8	12.4	1205	9	US-10-084-994-2
9	395.4	11.6	2700	9	US-10-084-994-5
10	175	5.1	1686	9	US-10-084-994-6
11	165.8	4.8	1578	12	US-10-105-929-10
12	156	4.6	1464	9	US-10-152-661-255
13	156	4.6	1464	9	US-09-866-050A-255
14	156	4.6	1627	9	US-10-152-661-438
15	156	4.6	1627	9	US-09-866-050A-438
16	154	4.5	1633	9	US-10-152-661-73
17	154	4.5	1633	9	US-09-866-050A-73
18	154	4.5	2569	9	US-09-796-753-115
19	134.4	3.9	393	10	US-09-833-381-1076

20	129.6	3.8	1939	9	US-10-125-459-2	Sequence 2, Appl1
21	129.6	3.8	1939	9	US-10-067-761-4	Sequence 4, Appl1
22	129.6	3.8	1939	10	US-09-804-156-4	Sequence 2, Appl1
23	129.6	3.8	1939	10	US-09-946-633-2	Sequence 133, App
24	89.2	2.6	3408	9	US-09-870-759-133	Sequence 1279, Ap
25	89.2	2.6	3412	9	US-09-870-759-133	Sequence 1279, Ap
26	85.8	2.5	8317	9	US-10-091-504-1279	Sequence 1279, Ap
27	85.8	2.5	8317	10	US-09-764-869-1279	Sequence 19, Appl
28	82.8	2.4	2598	9	US-09-842-758-19	Sequence 17, Appl
29	82.8	2.4	2804	9	US-09-842-758-17	Sequence 44, Appl
30	64.2	1.9	540	9	US-10-073-865-44	Sequence 215, App
31	64.2	1.9	540	10	US-09-764-853-215	Sequence 61, Appl
32	64.2	1.9	540	10	US-09-764-858-61	Sequence 148, App
33	54	1.6	3914	12	US-10-044-090-148	Sequence 18, Appl
34	47.4	1.4	2005	10	US-09-986-552-18	Sequence 18, Appl
35	47.4	1.4	2005	10	US-09-986-552-18	Sequence 149, App
36	43.4	1.3	2739	9	US-10-091-504-1280	Sequence 1280, Ap
37	43.4	1.3	2739	9	US-09-764-869-1280	Sequence 150, App
38	41.2	1.2	602	9	US-10-073-865-149	Sequence 925, App
39	41.2	1.2	602	9	US-10-073-865-150	Sequence 926, App
40	41.2	1.2	602	10	US-09-764-853-925	Sequence 305, App
41	41.2	1.2	602	10	US-09-764-853-926	Sequence 306, App
42	41.2	1.2	602	10	US-09-764-898-305	Sequence 391, App
43	41.2	1.2	602	10	US-09-764-898-306	Sequence 392, App
44	40.4	1.2	667	10	US-09-867-550-1911	
45	40.2	1.2	756	9	US-10-123-155-392	

ALIGNMENTS

Result 1
US-10-092-390-1
Sequence 1, Application US/10092390
Publication No. US20030013865A1
GENERAL INFORMATION:
APPLICANT: Yu, Xuanchuan
TITLE OF INVENTION: NO. US20030013865A1 Human BCF-Family Proteins and Polynucleo
FILE REFERENCE: LEX-0317-USA
CURRENT APPLICATION NUMBER: US/10/092,390
CURRENT FILING DATE: 2002-03-06
PRIOR APPLICATION NUMBER: US 60/275,013
PRIOR FILING DATE: 2001-03-12
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 1
LENGTH: 3423
TYPE: DNA
ORGANISM: homo sapiens
US-10-092-390-1

Query Match
Best Local Similarity 100.0%; Score 3423; DB 9; Length 3423;
Matches 3423; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCTATTCTTTTGAACATGCTGACCTTATTTGTTATTTGTTATGACCTGATG 60
DB 1 ATGCTATTCTTTTGAACATGCTGACCTTATTTGTTATTTGTTATGACCTGATG 60
QY 61 GGGACAGCATCACCCTGTAATCTGTAAGCCCTAATGCTGTAAGCCAGGAAAGCTAC 120
DB 61 GGGACAGCATCACCCTGTAATCTGTAAGCCCTAATGCTGTAAGCCAGGAAAGCTAC 120
QY 121 TCAGTACTGTCAGAGATCATACCATCCATCCCTTTGATCAATTTACTACAGAGCTGC 180
DB 121 TCAGTACTGTCAGAGATCATACCATCCATCCCTTTGATCAATTTACTACAGAGCTGC 180
QY 181 ACTGACATTTCTAACTGTTTAAATGACAGCGGACAGAGTACTGATGACAGCTAT 240
DB 181 ACTGACATTTCTAACTGTTTAAATGACAGCGGACAGAGTACTGATGACAGCTAT 240
QY 241 CACATCGGGAGAGACATATATAGGCGCAAGTCTCATGTTGTCCTGATTTATGAA 300

|||||
Db 241 CGACATGGGAGAGACTATGATAGCGCAGCTCAGTGTCTGCTGATTTATGAA 300
Qy 301 AGCGGGGAAATGTGTCCCCCACTGTGTATTAATGTGTCCATGGTGGCTGATTTGCT 360
Db 301 AGCGGGGAAATGTGTCCCCCACTGTGTATTAATGTGTCCATGGTGGCTGATTTGCT 360
Qy 361 CCAAAACCTGTCACTGTGAGCCGTGGGTGGGAGGAGCAACTGTCTCAGTGTGCGAT 420
Db 361 CCAAAACCTGTCACTGTGAGCCGTGGGTGGGAGGAGCAACTGTCTCAGTGTGCGAT 420
Qy 421 GGTATATCTTGGGTCCCCCACTGTGACCAAGCGGTGTCCAGTGTCAAAAATGGGGCTGTGTC 480
Db 421 GGTATATCTTGGGTCCCCCACTGTGACCAAGCGGTGTCCAGTGTCAAAAATGGGGCTGTGTC 480
Qy 481 AACCCATCAACCGGGGCTGTGCACTGTGTCTCGGGCTTCCGGGGCTGGCGCTGGAGAGAC 540
Db 481 AACCCATCAACCGGGGCTGTGCACTGTGTCTCGGGCTTCCGGGGCTGGCGCTGGAGAGAC 540
Qy 541 CGCTGTGAGCAGGCGACCTATGTATACGACTGTCTAGAGATGCCAGTGGCAGAAATGGA 600
Db 541 CGCTGTGAGCAGGCGACCTATGTATACGACTGTCTAGAGATGCCAGTGGCAGAAATGGA 600
Qy 601 GCCACCTGCGACACGTCACGCGGGGAAATGCCGTGCCACAGGATACACCGGAGCCTTC 660
Db 601 GCCACCTGCGACACGTCACGCGGGGAAATGCCGTGCCACAGGATACACCGGAGCCTTC 660
Qy 661 TGTGAGATCTTTTCTCTCTGTGTAAACATGTGTCCACAGTGTGAGCAGAGATGCCCTTGT 720
Db 661 TGTGAGATCTTTTCTCTCTGTGTAAACATGTGTCCACAGTGTGAGCAGAGATGCCCTTGT 720
Qy 721 CAAAATGAGAGAGTGTGTCTATCAGTCAGTCAGAGATGTCTTCCCTTCTGGCTGTGATG 780
Db 721 CAAAATGAGAGAGTGTGTCTATCAGTCAGTCAGAGATGTCTTCCCTTCTGGCTGTGATG 780
Qy 781 GGCACAGTGTGTGTGACGCTTGCCTCCGAGGGGTGGCTTGGAAAGAACTGTGCCAAGAA 840
Db 781 GGCACAGTGTGTGTGACGCTTGCCTCCGAGGGGTGGCTTGGAAAGAACTGTGCCAAGAA 840
Qy 841 TGCCAGTGGCCATTAATGAGAGGAGCGTGTATGCTCCACAGGCGCAATGTCTATGCAATGCA 900
Db 841 TGCCAGTGGCCATTAATGAGAGGAGCGTGTATGCTCCACAGGCGCAATGTCTATGCAATGCA 900
Qy 901 GGAATACACAGGAGGAGCGGTGTGACAGATGATGTCTGTGGAGCTATGGCGTTCTCTGT 960
Db 901 GGAATACACAGGAGGAGCGGTGTGACAGATGATGTCTGTGGAGCTATGGCGTTCTCTGT 960
Qy 961 GCTGAGACCTGTCCAGTGTGTCAACGAGGAGGAGTGTACCAAGTGTGAGGAGGAGCTGCTC 1020
Db 961 GCTGAGACCTGTCCAGTGTGTCAACGAGGAGGAGTGTACCAAGTGTGAGGAGGAGCTGCTC 1020
Qy 1021 TGTGAGACGAGGCTTGTGTGGGAGCGGTGTGCAAGACAGCGCTGTGTCTGAGAGGCTGTAC 1080
Db 1021 TGTGAGACGAGGCTTGTGTGGGAGCGGTGTGCAAGACAGCGCTGTGTCTGAGAGGCTGTAC 1080
Qy 1081 GGCATCAAAATGTGCAAAAGCGTGTCTGCTGCACTTGGAAAACTCATGTGTGTACACCC 1140
Db 1081 GGCATCAAAATGTGCAAAAGCGTGTCTGCTGCACTTGGAAAACTCATGTGTGTGTACACCC 1140
Qy 1141 ATGTCTGAGAGTGTGTGTGCAAGCGGGGTGTGAGAGCTCTAGTGAATGAGACATGT 1200
Db 1141 ATGTCTGAGAGTGTGTGTGCAAGCGGGGTGTGAGAGCTCTAGTGAATGAGACATGT 1200
Qy 1201 TGTCTGATGATTCACGAGGAGCTTGTGCAAGCAGATGTGACAGTGTGCAAAATGGGGAGAC 1260
Db 1201 TGTCTGATGATTCACGAGGAGCTTGTGCAAGCAGATGTGACAGTGTGCAAAATGGGGAGAC 1260
Qy 1261 TGTGACAGTGTGATGCAAAAGTGTGACAGTGTGACAGTGTGCAAAAGTGTGATGCTCT 1320
Db 1261 TGTGACAGTGTGATGCAAAAGTGTGACAGTGTGACAGTGTGCAAAAGTGTGATGCTCT 1320
Qy 1321 ACCCATGCGCTTGGGAACTATGGGATTAACGTGTCTGTGCTGTGCTGTGAAAAAT 1380
Db 1321 ACCCATGCGCTTGGGAACTATGGGATTAACGTGTCTGTGCTGTGCTGTGAAAAAT 1380

Db 1321 ACCCATGCGCTTGGGAACTATGGGATTAACGTGTCTGTGCTGTGCTGTGAAAAAT 1380
Qy 1381 GATCAGCTGTGTCTCTCTGTGTGAGAGGGGTCTTGTACTGTGAAGGAGCGGTGGACAGGGGT 1440
Db 1381 GATCAGCTGTGTCTCTCTGTGTGAGAGGGGTCTTGTACTGTGTGAAGGAGCGGTGGACAGGGGT 1440
Qy 1441 GACTGTCTCATCAGATGTCCAGTGTGAGCAGATGGGGCTTGTGTAACTTAACATGTCAG 1500
Db 1441 GACTGTCTCATCAGATGTCCAGTGTGAGCAGATGGGGCTTGTGTAACTTAACATGTCAG 1500
Qy 1501 TGTCTCAACGAGGAGGAGCTGTGCAACACCTGTGAGAGGAGCTGTGACAGTGTGACAGTGTG 1560
Db 1501 TGTCTCAACGAGGAGGAGCTGTGCAACACCTGTGAGAGGAGCTGTGACAGTGTGACAGTGTG 1560
Qy 1561 CGCGGGGAGAAATGCGAACTTCTCTGCGAGGATGGCAGTACGGGCTGAATGTGTGAG 1620
Db 1561 CGCGGGGAGAAATGCGAACTTCTCTGCGAGGATGGCAGTACGGGCTGAATGTGTGAG 1620
Qy 1621 CGCTGCACTGTGACAGCAGCAGATGGCTGCCACCTTACACAGGGGCTATGCGCTGTGCTC 1680
Db 1621 CGCTGCACTGTGACAGCAGCAGATGGCTGCCACCTTACACAGGGGCTATGCGCTGTGCTC 1680
Qy 1681 CCGGATGTGTGAGTGTGCACTGTGACAGGAGTGTGTGAGGAGGAGGAGGAGGAGGAGGAGG 1740
Db 1681 CCGGATGTGTGAGTGTGCACTGTGACAGGAGTGTGTGAGGAGGAGGAGGAGGAGGAGGAGG 1740
Qy 1741 TGTCTCTGCTGCTGTCTGTGTAAATGGGGCTTGTATGCTTCCCTGTATGTATGTCATGTC 1800
Db 1741 TGTCTCTGCTGCTGTCTGTGTAAATGGGGCTTGTATGCTTCCCTGTATGTATGTCATGTC 1800
Qy 1801 GAGTGTGACAGGAGCTTCCAGGAGCAGCTTGTGTAGAGATGTGTCTTCCCTGTGTATAT 1860
Db 1801 GAGTGTGACAGGAGCTTCCAGGAGCAGCTTGTGTAGAGATGTGTCTTCCCTGTGTATAT 1860
Qy 1861 GGGGATGCTGTGACAGCAGCAGATGGCCACAGTGTGTACAGCAGGAGGAGGAGGAGGAGGAG 1920
Db 1861 GGGGATGCTGTGACAGCAGCAGATGGCCACAGTGTGTACAGCAGGAGGAGGAGGAGGAGGAG 1920
Qy 1921 ATCAGCGGCTGTGTGACTGTGTGCTGTGTACAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1980
Db 1921 ATCAGCGGCTGTGTGACTGTGTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1980
Qy 1981 CCCAGTGGCAGATTTGGGAAAACTGTGAGAGATTTGTATGCTGTGTGTGTGTGTGTGTGTGT 2040
Db 1981 CCCAGTGGCAGATTTGGGAAAACTGTGAGAGATTTGTATGCTGTGTGTGTGTGTGTGTGTGT 2040
Qy 2041 TGTAAACCCATTTGACATCTTGTCACTGTGTATGCCGGTGTGTGTGTGTGTGTGTGTGT 2100
Db 2041 TGTAAACCCATTTGACATCTTGTCACTGTGTATGCCGGTGTGTGTGTGTGTGTGTGTGT 2100
Qy 2101 CAACCATGTCCACTGTCCACTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2160
Db 2101 CAACCATGTCCACTGTCCACTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2160
Qy 2161 GAGCTTGTGTGACAGGCTGTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2220
Db 2161 GAGCTTGTGTGACAGGCTGTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2220
Qy 2221 TACTGCACTAGAGATGTCTCTAGGGTTTATGAAAAAGTGTGTGCACTGTGTGTGTGTGTGT 2280
Db 2221 TACTGCACTAGAGATGTCTCTAGGGTTTATGAAAAAGTGTGTGCACTGTGTGTGTGTGTGT 2280
Qy 2281 TGTCAAAAGGAGTGTGAGTGTGAGCAGATTTGTGGGAGTGTGTGTGTGTGTGTGTGTGTGT 2340
Db 2281 TGTCAAAAGGAGTGTGAGTGTGAGCAGATTTGTGGGAGTGTGTGTGTGTGTGTGTGTGTGT 2340
Qy 2341 ATGGAGGAGCAGTGTGAGCAGAAATGCGCTTCAAGAACTATGTGTGTGTGTGTGTGTGTGT 2400
Db 2341 ATGGAGGAGCAGTGTGAGCAGAAATGCGCTTCAAGAACTATGTGTGTGTGTGTGTGTGTGT 2400
Qy 2401 ATATGTATGT 2460
Db 2401 ATATGTATGT 2460


```

PRIOR FILING DATE: 1998-12-30
PRIOR APPLICATION NUMBER: 09/259,388
PRIOR FILING DATE: 1999-02-26
PRIOR APPLICATION NUMBER: 60/122,458
PRIOR FILING DATE: 1999-03-01
PRIOR APPLICATION NUMBER: 09/312,359
PRIOR FILING DATE: 1999-05-14
PRIOR APPLICATION NUMBER: 09/336,536
PRIOR FILING DATE: 1999-06-18
PRIOR APPLICATION NUMBER: 09/342,687
PRIOR FILING DATE: 1999-06-29
PRIOR APPLICATION NUMBER: 09/345,464
PRIOR FILING DATE: 1999-06-30
PRIOR APPLICATION NUMBER: 09/365,164
PRIOR FILING DATE: 1999-07-30
PRIOR APPLICATION NUMBER: 09/399,723
PRIOR FILING DATE: 1999-09-20
PRIOR APPLICATION NUMBER: 09/409,634
PRIOR FILING DATE: 1999-09-30
PRIOR APPLICATION NUMBER: 09/471,179
PRIOR FILING DATE: 1999-12-23
PRIOR APPLICATION NUMBER: 09/474,071
PRIOR FILING DATE: 1999-12-29
PRIOR APPLICATION NUMBER: 09/474,072
PRIOR FILING DATE: 1999-12-29
PRIOR APPLICATION NUMBER: 09/514,010
PRIOR FILING DATE: 2000-02-25
PRIOR APPLICATION NUMBER: 09/516,745
PRIOR FILING DATE: 2000-03-01
PRIOR APPLICATION NUMBER: 09/572,002
PRIOR FILING DATE: 2000-05-14
PRIOR APPLICATION NUMBER: 09/597,993
PRIOR FILING DATE: 2000-06-19
PRIOR APPLICATION NUMBER: 09/599,596
PRIOR FILING DATE: 2000-06-22
PRIOR APPLICATION NUMBER: 09/630,334
PRIOR FILING DATE: 2000-07-31
PRIOR APPLICATION NUMBER: 09/606,565
PRIOR FILING DATE: 2000-06-29
PRIOR APPLICATION NUMBER: 09/606,317
PRIOR FILING DATE: 2000-06-29
PRIOR APPLICATION NUMBER: 09/665,666
PRIOR FILING DATE: 2000-09-20
PRIOR APPLICATION NUMBER: 09/677,751
PRIOR FILING DATE: 2000-09-30
NUMBER OF SEQ ID NOS: 162
SEQ ID NO 123
LENGTH: 3567
TYPE: DNA
ORGANISM: Rattus sp.
FEATURE:
NAME/KEY: CDS
LOCATION: (925)...(2832)
US-09-796-753-123

```

```

Query Match      17.8% Score 608.4; DB 9: Length 3567;
Best Local Similarity 56.7%; Pred. No. 58e-188;
Matches 126; Conservative 0; Mismatches 901; Indels 37; Gaps 4;

```

```

QY 67 GCATCACTCTGATCTTGAAGACCTTAATGTGTAGCCACTGGGAAGACTACTAGTG 126
DB 236 GGTGAACACTCAATCCATGATGCCATGTGTACTCTGGGAAGCTTACACAG 295
QY 127 ACCTGCAAGAGTCATACCACTCCCTTT--GATCAATTACTACAGAGCTGCACT 183
DB 296 ACCACTAAGAGTCCACCTTCGCTTCAGCTCCGCCACGAGTCTGCGACAG 355
QY 184 GACATCTTAACCTGTTTAATGACGCGACAGAGTCAAGTCAAGGAGAGCTTACGA 243
DB 356 CCTTGGGAAGACCCACACCTGCGCTCAGACCTACGTTGTCTACCGGAGCTGTACCG 415
QY 244 CATGGGAGAGACTATGATAGCGCAAGTCTCACTGTTGCTCGATTATTAAGAAC 303

```

```

DB 416 CAGTGTGTGAAGATGAGTCCCGCCACGCTGACGTGCTGTTGGGGTACTACAGAGC 475
QY 304 GGGGAATGTGTCTCCCACTGTCTGATTAATGTCTCAATGTGCTGTATGCTCA 363
DB 476 AGTGAGCCGTGTCTCCCACTGTCTGCGAGAGTGTCTCAGAGTGTGCTGTCT 535
QY 364 AACACCTGTACAGTGTAGCTGTGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 423
DB 536 AATCGGTGCCAGTGTGACACAGGCTGTGGGGGAGGAGGAGGAGGAGGAGGAGG 595
QY 424 GATCACTGGGGTCCCACTGACACAGCCGCTGACAGTGAATAAATGGGCTGTGCAAC 483
DB 596 GGAATGGGGAGACACAGTGTGACAGGCTGTCTGTGGCAACAGCAATGCTGTGAT 655
QY 484 CCCATACCGGGGCTTGCACCTGTCTGTGGGCTTCCGGGCTTCCGAGGAGCCG 543
DB 656 CCCAGAGGGGGTGTGTGTTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGT 715
QY 544 TGTGACAGGGGACATGATGTAAAGACTGTGATCAAGATGCCAGTCCAGAAATGGAGC 603
DB 716 TGCCCGGATGGCCATATGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGT 772
QY 604 ACCTGGACACGTCACGGGGGAGATGCGCTGCCACAGATACCGGAGCTTCTGT 663
DB 773 TCTGTGATGCCCCCGGATGAGGCTGTCTGTGCTTGTGCTTGTGCTTGTGCTTGT 832
QY 664 GAGATCTTGTCTCTCTGTGTAAACATGTGTGCTTGTGAGAGAGATGCTTGTCA 723
DB 833 CTGATGCTTCTTCTGCCCCAGAAC-----TTATGCTTGTCA 870
QY 724 AATGAGAGATGTGTATCATGACGTCACGTAAGTGTGCTTGTGCTTGTGCTTGTGCT 783
DB 871 AATGAGAGTGTCTCTAAGGCTGTCAAGCTCTGTGAGTGTGCTTGTGCTTGTGCT 930
QY 784 ACAGTGTGTGTCAGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGT 843
DB 931 GTGATGTGTCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCT 990
QY 844 CAGTGTGTAAAGAGGAGGAGTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCT 903
DB 991 CTTTGCACAAATGTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCT 1050
QY 904 TACACAGGGGAACGCTGAGAGTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCT 963
DB 1051 TATATGGGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1110
QY 964 GAGACCTGCACTGTGTCAAGGAGGAGGAGTGTACACGCTGAGCGGCGACTCTGT 1023
DB 1111 GAGACCTGTGACTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1170
QY 1024 GAAGAGGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1083
DB 1171 GAACATGGCTTCAACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1230
QY 1084 ATCAATGTGACAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1143
DB 1231 CTGAGTGTGCAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1290
QY 1144 TCTGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1203
DB 1291 CAGGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1350
QY 1204 CTTGATTTAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1263
DB 1351 CAGGAGACGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1410
QY 1264 GACAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1323
DB 1411 CTGCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1470
QY 1324 CCATGCTTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1383
DB 1471 CTTTGTGCAACTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1530

```



```
635 GCCACAGAGATACACCGAGCTTCTGTAGAGATCTTGTCTCTCTGTAACATGTC 694
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 888 GCCCGGACAGAGAACTGGGACAGTGTGAGTGTCTGCTTCCAGGACATTTCTGCT 947
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 695 CACAGTGTAGAGAGATGCTTGTCAAAATGAGAGAGTGTGTATCATCACTGAG 754
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 948 TCTTGTGCCCCAGACACCCATCTTGCCTGCAAAATGAGAGTGTCTTCCAAACCCACAGGGCT 1007
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 755 AATGCTTTGGCTTGTGCTGTAGAGGACAGTGTGTGTGTAGAGCTTGTGCCCCGAGGGTC 814
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1008 CCTGCACTGCCCCCTTGTGATGTGAGTGTGATGCTGCTTGTGCTGCTGCTGCTGCTGCTGCT 1067
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 815 GCTTTGAAAGAACTGTTCCCAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 874
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1068 TTTACAGAGACCAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1127
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 875 CCACAGGCGCAATGTCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 934
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1128 TCACTGAGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1187
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 935 CTGTTGGAGCTATGAGCGCTTCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 994
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1188 CGGTGGGCGCTTGTGAGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1247
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 995 GTTACACAGTGAAGCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1054
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1248 GCTTCCGCGCAAGCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1307
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1055 CACGCTGTGTCTGAGGGGCTGTGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1114
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1308 ATGCGCTGTGCGCCAGCGCTTCTGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1367
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1115 TGAAGAACTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1174
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1368 GGGAGACAGCTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1427
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1175 CAGGACTTACTGTATGAGACATGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1234
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1428 CGGGCTCTCACTGCAACGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1487
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1235 TCTGACGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1294
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1488 ACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1547
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1295 CAGGATTCAGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1354
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1548 CGGGTTCACGCGGCTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1607
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1355 GTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1414
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1608 GTTCTGACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1667
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1415 CTTGCAAGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1474
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1668 TCTGCAAGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1727
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1475 GCTTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1534
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1728 GCTTGAAGTTCAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1787
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1535 GGAAGCTGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1594
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1788 GAGCTGTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1847
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1595 GCACTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1654
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1848 GCACTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1907
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1655 CTACACAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1714
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1908 CTGTTCACTGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1967
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1715 GTGCTGAGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1774
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
Db 1968 GCCCTGAGGCGCTTATGAGGAGTCACTGATGACACACCTGACCTGCAAGATGGGGCA 2027
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1775 CATGCTCCCTATGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1834
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2028 CCTGCTCTCTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2087
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1835 AGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1894
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2088 AGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2141
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1895 TTTACAGAGACGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1954
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2142 CTACACAGCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2198
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1955 CAGGCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2014
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2199 CAGGCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2258
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 2015 TTTGATACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2074
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2255 CTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2318
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 2075 CCGGTTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2134
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2319 TAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2378
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 2135 GCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2194
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2379 GCTCCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2438
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 2195 AATGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2207
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2439 TATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2451
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 7
US-09-833-381-1910/c
; Sequence 1910, Application US/09833381
; Patent No. US20020132090A1
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: NO. US20020132090A1el Nucleic Acid and Protein Homologs
; FILE REFERENCE: 5800-119
; CURRENT APPLICATION NUMBER: US/09/833,381
; PRIOR FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 09/516,448
; NUMBER OF SEQ ID NOS: 2050
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO: 1910
; LENGTH: 5197
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)..(5197)
; OTHER INFORMATION: n - A,T,C or G
US-09-833-381-1910

Query Match 15.3%; Score 524.6; DB 10; Length 5197;
Best Local Similarity 58.0%; Pred. No. 2.3e-160;
Matches 972; Conservative 0; Mismatches 694; Indels 11; Gaps 2;
```


Db	132	CTCCCACTAATATGGCTCTCTGTACTGTCAAGAGAGGGGTGGCAGGGCTTGACTGCACCT	241
Oy	1452	CAGATGTCCAGTGGACACATGGGGCTTTGGCTGTAACTTAACATGGCACTGCTCAAGG	1511
Db	242	GCAATGTCCAGATGGGAGCGTGGGGCCTGAACTGCACAGAGGTGCACCTGTGGCAATGG	301
Oy	1512	GGGAGCTCTGCACACCTCGACGGAGACTTGCACGTGTGCACCTGGATGGCGGGGAGAA	1577
Db	302	GGCAGCTCGACGCCCATAGACGGCTCTCTGCTCACTCCGTGGCTGTGGAGAGAC	361
Oy	1572	ATGGGAATCTCCGTCAGAGATGGCAGTACAGTACGGGGTGAACCTGGTGAGACGGTGGAGAC	1633
Db	362	CTGTGAGCTGCTTGGCCCGGATGGCAATTTGGGCTGAACTGGAGTAACTGTGACTG	421
Oy	1632	CAGCCACGACAGATGGCTGTCACACCTTACACAGGGGCATTTCCGCTGCTCCCGGATGGTC	1691
Db	422	CAGCAATCTGATGGATGTGACCCCGTCACAGGCCACTGTGCTGTGGCCGAGTGAAC	481
Oy	1652	AGGTGTCACTGTGACAGCGCTGTGTGCTGAGGAGCGTGGGGGCCCAACCTGCTCCCTGCC	1751
Db	482	AGGCATCCGCTGTACAGACAGCTGTCCACCTGCGCCGCTGGGGCCCAACTGCTGTGTC	541
Oy	1752	CTGCACTGTAAATATGGGGCTTCATGCTCCCTGATGATGGCATGTGGAGTGTGCAC	1811
Db	542	CTTGAGCTGTAGATATGAGAGCTCTGTGCTCCCGAGAGATGGGAGCTGGAGTGTGGCC	601
Oy	1812	AGGCTTCCGAGGACACCACTTGTGAGAGAGATCTCTCCCTGGGTTTATGGGCAATCGCG	1872
Db	602	TGGCTTCGAGAGACCTTATGTCCAGAGAAATCTGCCCCCTGGGTTTATGGCCACAGGCTG	661
Oy	1872	CAGCCAGACATGCCACAGTGTGGTTACAGACAGACGGGGCCTGTGCACACATCACCGGCT	1931
Db	662	CGCCACGACATGCCCTCTGCGGTCACAGACAGAGCCCTGTGCACACATCAGGGGCAAT	721
Oy	1932	GTGTGACTGCTTGGCTGGCTTACAGAGCGCCCTCTGTCAATGAAG	1975
Db	722	CTGTGATGCTCTCCAGAGATCTCTGGAGCTCTCTCAACACAG	765

```

RESULT 9
US-10-084-994-5
; Sequence 5, Application US/10084994
; Publication No. US20030023070A1
; GENERAL INFORMATION:
; APPLICANT: NI et al.
; TITLE OF INVENTION: Attractin-Like Polynucleotides, Polypeptides, and Antibodies
; FILE REFERENCE: PTO11P1C1
; CURRENT APPLICATION NUMBER: US/10/084,994
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/790,621
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: PCT/US00/23663
; PRIOR FILING DATE: 2000-08-29
; PRIOR APPLICATION NUMBER: US 60/151,348
; PRIOR FILING DATE: 1999-08-30
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 5
; LENGTH: 2700
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-084-994-5

```

	Query Match	Similarity	Score	395.4	DB	9	Length	2700
	Best Local	Similarity	52.0%	Pred. No.	3.5e-18			
	Matches	1019	Conservative	0	Mismatches	921	Indels	21
								Gaps
								5
Oy	407	CCAGTGCCTGCATGTCATCTAGTGGGGTCCCATCGCAGCCGCGGGTCCAGTGC	AAA	466				
Db	15	CCCACGCGTCCACGACAGGCTGTATGTGTCACAGTCCAGACAAAGTCTTTTGCCA	74					
Oy	467	ATGGGGCTCTGTGCACCCCATCACCGGGGCTTGCACCTGTCCTCGGGCTTC	526					

[illegible]


```

QY 1601 ACGGCTGAACTGCTGAGCGCTGCGACCTGCACGACGAGATGCTGCCACCTACCA 1660
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1206 ATGGGCGAGGCTGTGAACAGCTGTGGGTCTCAACGGGGGCTCCTGTATGATCGGCA 1265
QY 1661 CGGGGCAATGGCGCTGCTCCCGGATGCTGAGTGTCCACTGTGACAGCGTGTGTCTG 1720
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1266 CGGGGCGCTGCGCTGCGCCCACTGGTTCCTCGGGAGGAGTGCACCTGCTGCCG 1335
QY 1721 AGGAGCGCTGGGGCCCGCCCACTGCTCCCTGCGCTCTACTGTAAATAATGGGCTCATGCT 1780
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1326 AGGGCGCTGTGGGCCCACTGCGACGAGTGTGGGTGGGGCAGGGGGCGGCTGCG 1385
QY 1781 CCGCTGATGATGCAATGTGGAGTGTGCACAGGCTTCCAGGACCACTTGTCAAGGA 1840
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1386 ACCCTGTGACGCGGACCTGCTGCGCCCGGGGAGAGCGCGCTGCTGAGGAG 1445
QY 1841 TCTGCTCCCGGTTTTATGGGCAATGCGTGCACGCCAGACATGCCACAGTGCCTTACA 1900
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1446 GCTGCCCGCCAGAACGGGTTTTGGCGTGGCGTGCAGACACCTGCTCTCGAGAAAT--- 1501
QY 1901 GCAAGCGGCGCTGCGCACACATCACCGGCTGTGTGACTCTTGGCTTGCACAGCG 1960
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1502 -GGGGGCTGTGCGCAGCGCAGCAAGGCAAGCTGCTCTGTGGCTGTGGCTGAGCGGG 1559
QY 1961 CCGCTGCAATGAAATGTGTCCAGTGCAGATTTGGGAAAACTGTGCAGGAATTTGTA 2020
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1560 GGCAGCTCGAGCTGCGCTGTCCCGCTGGGCGCTACGGAGCGCGCTGCATCTGGAGTGT 1619
QY 2021 CCTGACCAACAGGAAGCACTGTAAACCCATGTGACAGATCTGTGCACTGTATCCCGGTT 2080
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1620 CTTGCCCAACAGACAGCTGTGAGCTGTGCACGCGGCACTGCGCTGGCGGCCGCT 1679
QY 2081 GGATTTGCACTGCTCTCAACATGTCCACCTGCCCACTGGGGCCCAACTGCATCTCC 2140
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1680 TCTATGCGCAGGCGCTGCGAGACCCCTGTGCCCTGTGCTTCCAGGGGCTGGCGCAGG 1739
QY 2141 ACAGTTCAACTGCGCATTAATGAGACTTTCGACGCGCTTACGATGGGAATGTAATGCA 2200
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1740 GATTGTCTGTGTCAACATGTGACCCCTGCGACCCCACTAGTGGCGGCTCTGCGC 1799
QY 2201 CTTCTGCTGGAGCGGCTGTACTGACCTCAGAGATGCTCTTATGTTATGGAAG 2260
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1800 CTTGCGGCTTCCAGCGGCACTTCTGTAGAGGGGCTGTGAGCCAGTTTATGGAGAG 1859
QY 2261 ATTGTGCACTGATATGCCAATGTCAAAACGAGCTGCACTGCAGCACATTTCTGGGAGT 2320
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1860 GTGCGCACAGCGCTGTGACTGTGAGGGGGGAGACCTGTGACCCCTGTACCGGCTCT 1919
QY 2321 GTACTTGGCGCACTGATTTATGGGAGCGGCACTGTGAGCAG 2361
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1920 GCCTTTGCCACAGGGCGCTCAGAGGACCACTGTAACTGT 1960

```

```

RESULT 10
US-10-084-994-6
; Sequence 6, Application US/10084994
; Publication No. US20030023070A1
; GENERAL INFORMATION:
; APPLICANT: NI et al.
; TITLE OF INVENTION: Attraction-Like Polynucleotides, Polypeptides, and Antibodies
; FILE REFERENCE: PT011P101
; CURRENT APPLICATION NUMBER: US/10/084,994
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/790,621
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: PCT/US00/23663
; PRIOR FILING DATE: 2000-08-29
; PRIOR APPLICATION NUMBER: US 60/151,348
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 6

```

```

; LENGTH: 1686
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-084-994-6
Query Match 5.18; Score 175; DB 9; Length 1686;
Best Local Similarity 51.6%; Pred. No. 4,8e-46;
Matches 477; Conservative 0; Mismatches 440; Indels 8; Gaps 3;
QY 1165 CCGGGCTGTGAGAGCTCTACTGTAAATGAGACATGTTCTCTCGATTTCTAGGGGAAGCT 1224
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 15 CTTGCTTCACTGTGCTCCGCTGCGAGAGATGTCCCGGGCGGTATGGGCCAGGC 74
QY 1225 TGGCAGAGATCTGCACCTGCCAATAATGGGGCAGAGCTGGA-CAGTGTGATGGAAGTG 1283
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 75 TGTGAACAGCTGTGTGTGCTCTCAAGGGGGCTCCTGTATGCGGCCAGGGGGGCTG 134
QY 1284 CACCTGTGCCCGCAGATTCAAAGAAATGACTGCT-CTACCCATGCGCTCTGGAACT 1342
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 135 CCGCTGCCCGCAGTGGGTTCTCGGGAGAGGACTGCACCTGATCTCCAGAGCGCGCT 1394
QY 1343 ATGGGATTAATCTGTCTCTCGCTGTGGGCTGTAAATAATGAGTGTGCTCTGTGTG 1402
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 195 TCGGCCCACTGCACCCACAGTGTGTGGGTGTGGGCGAGGGGGCGCTTGCAGCCCTGTGA 254
QY 1403 ACGGCTTTTACTTGCMAAGGAGCGTGCACGCGGGTGTGACTGCTCCATGAGATGCCA 1462
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 255 CCGGACCTGCTCTGTGCGCCCGGGGAGAGCGCGGCTGCTGAGAGGAGGCTGCC 314
QY 1463 GTGGCAATGGGCTTTGGCTGTAACTTAATGCAATGCCAGTGTCTCAAGGGGGAGCTTGA 1522
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 315 AGAACCGGTTTGGCTGTGGGCTGTGAGACACCTGCTCTCAGAAATGGGGGCTGTGCC 374
QY 1523 ACAACCTGAGAGGAGCTGTGACAGTGTGCACCTGATGAGCGCGGGAGAAATGCGAACTTC 1582
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 375 ACGCCAGCAACGGCAGCTGTCTGTGTGGCTGTGGGCTGTGAGGGGGGCACTGTGAGCTGG 434
QY 1583 CTTGCCAGATGAGCAGTGTGAGGGCTGAACTGTCTGAGCGCTGCAGTGCAGGACGAGCAG 1642
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 435 CTTGCTCCCGCTGGGCGCTGAGAGAGCGCGCTGCATCTGAGTGTGCTCTGCACAAACA 494
QY 1643 ATGGCTGCCACCTTACACAGGGGCTATGCGCTGCCGCGGATGTGACAGTGTCCACT 1702
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 495 GCACTGTGAGCTGTGCACAGGGGCACTGCGCGGCGCCCGCTTCTATGAGCGAGGCT 554
QY 1703 GTACAGACGCTGTGCTGAGAGGAGCGTGGGGCCCAACTGCTCCCTGTGCTGTACTGTA 1762
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 555 GCGAGACACCTGTCTCCCTGTGCTTCCAGGGGCTGTGACAGGGGTTTGTCCGGTGTG 614
QY 1763 AAAATGGGCTTACGTCTCCCTGTGATGAGCATCTGCGAGTGTGCACAGGCTTCCGAG 1822
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 615 AACATGAGAGCCCTGCGACCCCATCATGATGGCGAGATGCTCTGCGCGGCTTCCAG 674
QY 1823 GCAACACTTGTGAGAGATGCTCTCCCTGTGTTTATGGGCAATGCTGTGACGCAACAT 1882
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 675 GCGCACTTCTGTGAGAGGGGCTGTGAGCCAGGTTTATGGAGGGCTGTACACAGGCT 734
QY 1883 GCCCAGAGTGCCTTACAGAGGAGCGGCGCTGCCACACATCACCGGCTGTGACTGCT 1942
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 735 G-----TGACTGTGAGGGGGGGGACCCGTGTATACCCCTGTACACGGTCTCTGCTTTGCC 788
QY 1943 TGCCTGCTTACAGAGCGCGCTTGTGCAATGATGTGTCCAGTGTGCAGATTTGGGAAAA 2002
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 789 CACACAGGGGCTCAGAGGACCACTGTAACTGATTTGCAAGAGGGCCAGTTTGGGCCCA 848
QY 2003 ACTGTGAGGAATTTTACTGACCAACCAACAGGAACCTGTAAACCATGTGACAGATCT 2062
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 849 GCTGACCTTGCACCTGTGATGTGGGGGTGTGAGCTGTGACGACCTGTACGTAGGCGAGT 908
QY 2063 GTGAGTGTAAACCGGTTGGATTTGG 2087
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 909 GTGACTGTGTGATGTGCTATATGGG 933

```



```

RESULT 11
US-10-105-929-10
; Sequence 10, Application US/10105929
; Patent No. US20020137142A1
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas A.
; APPLICANT: Goodheart, Andrew D.J.
; TITLE OF INVENTION: TANGO-71, TANGO-73, TANGO-74, TANGO-76, AND TANGO-83
; FILE REFERENCE: 09404/041001
; CURRENT APPLICATION NUMBER: US/10/105,929
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/130,491
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-08-07
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/058,108
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-05
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/054,961
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-06
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 10
; LENGTH: 1578
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(1578)
; OTHER INFORMATION: n = A,T,C or G
US-10-105-929-10

```

```

Query Match
Best Local Similarity 4.8%; Score 165.8; DB 12; Length 1578;
Matches 212; Conservative 0; Mismatches 3; Indels 5; Gaps 4;

```

```

QY 3209 CCACAGCAATG-TCATATGAGTTGAACCT-ACAGTGAAGTGTGT--CCAGGAGTATTC 3264
DB 5 CCACAGCAATGTTCTATGAGTTGAACCTACAGTGAAGTGTGTCCCAAGAGTATTC 64
QY 3265 AGCAATATGAGGCGCTCTCC-AGATCATATGACCTCCCAAGAGTCACTATCC 3323
DB 65 AGCAATATGAGGCGCTCTCTCCAGAGTCACTATGACCTCCCAAGAGTCACTATCC 124
QY 3324 TTGTCAATATGACCTCTGCACTCCAGAGTCACTATCCCTTAAGCAAGAGACAG 3383
DB 125 TTGTCAATATGACCTCTGCACTCCAGAGTCACTATCCCTTAAGCAAGAGACAG 184
QY 3384 TGAGGTAGCAGCAGCAAGCAGCAGCAGCAGCAGTGAATGA 3423
DB 185 TGAGGTAGCAGCAGCAGCAGCAGCAGCAGCAGTGAATGA 224

```

```

RESULT 12
US-10-152-661-255
; Sequence 255, Application US/10152661
; Publication No. US20030022835A1
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murlison, James G.
; APPLICANT: Kumble, Krishanand D.
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; FILE REFERENCE: 11000.1011c5
; CURRENT APPLICATION NUMBER: US/10/152,661
; CURRENT FILING DATE: 2002-05-20
; PRIOR APPLICATION NUMBER: 09/866,050
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 60/221,232
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: 60/206,650
; PRIOR FILING DATE: 2000-05-24

```

```

; PRIOR APPLICATION NUMBER: 09/312,283
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: PCT/NZ99/00051
; PRIOR FILING DATE: 1999-04-29
; PRIOR APPLICATION NUMBER: 09/188,930
; PRIOR FILING DATE: 1998-11-09
; PRIOR APPLICATION NUMBER: 09/069,726
; PRIOR FILING DATE: 1998-04-29
; NUMBER OF SEQ ID NOS: 725
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 255
; LENGTH: 1464
; TYPE: DNA
; ORGANISM: Mouse
US-10-152-661-255

```

```

Query Match
Best Local Similarity 4.6%; Score 156; DB 9; Length 1464;
Matches 419; Conservative 51.2%; Pred. No. 7.5e-40; Mismatches 390; Indels 9; Gaps 2;

```

```

QY 491 CCGGGGCTTGCCACTGTGCTGCGGGCTTCGGGGCTGGCGCTGCGAGACCGCTGAGC 550
DB 9 CCGGAGCTGCTACTGCTGCTGCGGGCTTCGGGGCTGCGAGACCGCTGCTGAGC 68
QY 551 AGGCACTATGAGTGAAGTCACTGTCATGAGATGCGCAGTGCAGAAATGAGCAGCTGCG 610
DB 69 AGGCTGCTTGGGGGAGTGGCCAGCTGTCGCGAGTGTGATGATGCGGAGGCGGATGTG 128
QY 611 ACCAGCTCAAGGGGAGTCCCGTGGCCACAGATATACCGAGGCTTGTGAGATG 670
DB 129 ACCAGCTGCGGGAGTCTGATGATGTCCTGCGGAGAGAGGAGGAGCATTGAGACGCG 188
QY 671 TTGTGCTCTGCTTAACATGTGTCACAGTGTGAGAGAGATGCTTGTGCAAAATGAG 730
DB 189 GCTGCTCCAGGAGCGGTTTGGCAAGGCTGTGTAACAAGTGTGCTCAGAAATGAGG 248
QY 731 GAGTGTGATCACTGCTGAGTGAATGCTTGTGCTTGTGCTGAGTGGCAGAGTGT 790
DB 249 GCTGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 308
QY 791 GTGTGAGCTTGGCCCGAGGCTGCTTGTGGAAGAAGTGTGCAAGATGCGAGTGC 850
DB 309 GTGAGCAGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 368
QY 851 ATATGAGGAGAGTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 910
DB 369 AGAACAATGAGCAGTGTGAGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 428
QY 911 GGGAGCGTGGCAGATGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 970
DB 429 GTCAAGCTTGTGAGAGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 488
QY 971 GCCAGTGTGCAACGAGGAGGAGTGTACCACTGAGCGGCTGCTGCTGCTGCTGCTG 1030
DB 489 GCGAGTGTGCAACGAGGAGGAGTGTACCACTGAGCGGCTGCTGCTGCTGCTGCTGCT 548
QY 1031 GCTTGTGCTGCGAGCGCTGCGAAGCAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1090
DB 549 GCTTGTGCTGCGAGCGCTGCGAAGGAGG--GTGCAAGCGAGGCTTGTGAGATGCT 605
QY 1091 GTGCAAAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1150
DB 606 GCTGAGAGAGTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 659
QY 1151 AGTGTGCTGCAAGCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1210
DB 660 TGTGCTTGTGCGCCAGGAGGCGGAGGAGCAACATGATGATGATGATGATGATGATG 719
QY 1211 TGTAGGGGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1270
DB 720 GCTTGTGCGCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 779
QY 1271 TGACTGAAAGTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1308

```

Db 780 TCAGTGGGACAGTCCACTGTGTGACACACTACACGGGA 817

RESULT 13

US-09-866-050A-255
Sequence 255, Application US/09866050A
Publication No. US20030040471A1

GENERAL INFORMATION:

APPLICANT: Watson, James D.
APPLICANT: Strachan, Lorna
APPLICANT: Sleeman, Matthew
APPLICANT: Onrust, Rene
APPLICANT: Murison, James G.
APPLICANT: Kumble, Krishanand D.
TITLE OF INVENTION: Compositions Isolated from Skin Cells
FILE REFERENCE: 11000.1011c4U
CURRENT APPLICATION NUMBER: US/09/866.050A
CURRENT FILING DATE: 2001-05-24
NUMBER OF SEQ ID NOS: 725
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 255
LENGTH: 1464
TYPE: DNA
ORGANISM: Mouse
US-09-866-050A-255

Query Match
Best Local Similarity 51.28; Pred. No. 7.5e-40;
Matches 419; Conservative 0; Mismatches 390; Indels 9; Gaps 2;

4.68; Score 156; DB 9; Length 1464;

Qy 491 CCGGGGCTTGCACATGTCTGCGGGGCTTCCGGGGCTGCGGAGACCGCTGTGAGC 550
Db 9 CGGAGCTGCTGACTAGCCCTGTGGGTTCTTGGGGCGAGCTGACCTTGTCTCCAC 68
Qy 551 AGGAGCTATGTGTACACATCTCTACAGATGCCAGTCCAGAAATGAGCCACTGCG 610
Db 69 AGGTCGCTTGTGGCCCGCCAGCTGTGCCACGTGTACATGCGGGGAGGGGCGCATGTG 128
Qy 611 ACCAGCTACGAGGGGAGATGCCCTGCCACAGATACACCGGACCTTCTGTGAGATC 670
Db 129 ACCAGTGTGCGGAGCTGTGATCTCTCTCCGGAGACGAGGAGCCATTTGTGAGCGG 188
Qy 671 TTTGTCTCTGTGTAAACATGTGTCCAGATGTGACAGAGATGCCCTTGTCAAAATGAG 730
Db 189 GCTGTCCCGACGAGCGTTTGGCAAGGCTGTGAACACAGTGTCTGTGAGAAATGAGG 248
Qy 731 GAGTGTGTATCACTCACTGTGAGAAATGTCTTGTGCTTGTGCTGTGATGGGCAAGTGT 790
Db 249 GCTGTGTATGTCTACCAATGTGACACTGTCTCTGCCCCCTGGGCTGTGATGGGCGCAC 308
Qy 791 GTGTACACCTTGTGCCCGAGGCTGTTGAAAGAACTGTTCACAAATGCCAGTCC 850
Db 309 GTGAGACAGCTGCGCTGTGCGGCTATGTGTGTGCTGTGCTGTGAGATGTTCTGTG 368
Qy 851 ATATGTGAGGAGCTGTGTATGTGTGACAGGAGCCAAATGTCAATGTGACAGTCAAGATACAG 910
Db 369 AGAACAAATGAGAGCTGTGAGCCACCTCGGGGCTTGTGCTGTGCTGTGCTGTGATG 428
Qy 911 GGGAGCTGTGAGAGTGTGTGTGTGAGACTATGTGAGCTGTGTGTGTGTGAGACT 970
Db 429 GTCAAGCTTGTGTGAACACACTGTGCTGCGGCTTGTCAATGTGTGTGTGTGAGAGATT 488
Qy 971 GCCAGTGTGTCAACGAGGAGAGTGTATACAGTGTGAGGCGGAGATGCTTGTGAAGAG 1030
Db 489 GCGAGTGTCAACAGGAGGCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 548
Qy 1031 GCTTGTGTGAGGAGCTGTGAGAGACAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1090
Db 549 GCTTGTGTGAGGAGCTGT 605
Qy 1091 GTGACAAACGAGT 1150

Db 606 GCTGTACACAGCTGT 659
Qy 1151 AGT 1210
Db 660 TGT 719
Qy 1211 TGT 1270
Db 720 GCTTGTGTGAGGAGCTGT 779
Qy 1271 TGACTGT 1308
Db 780 TCAGTGGGACAGTCCACTGTGTGACACACTACACGGGA 817

RESULT 14

US-10-152-661-438
Sequence 438, Application US/10152661
Publication No. US20030022835A1

GENERAL INFORMATION:

APPLICANT: Watson, James D.
APPLICANT: Strachan, Lorna
APPLICANT: Sleeman, Matthew
APPLICANT: Onrust, Rene
APPLICANT: Murison, James G.
APPLICANT: Kumble, Krishanand D.
TITLE OF INVENTION: Compositions Isolated from Skin Cells
FILE REFERENCE: 11000.1011c5
CURRENT APPLICATION NUMBER: US/10/152.661
CURRENT FILING DATE: 2002-05-20
PRIOR APPLICATION NUMBER: 09/866,050
PRIOR FILING DATE: 2001-05-24
PRIOR APPLICATION NUMBER: 60/221,232
PRIOR FILING DATE: 2000-07-25
PRIOR APPLICATION NUMBER: 60/206,650
PRIOR FILING DATE: 2000-05-24
PRIOR APPLICATION NUMBER: 09/312,283
PRIOR FILING DATE: 1999-05-14
PRIOR APPLICATION NUMBER: PCT/NZ99/00051
PRIOR FILING DATE: 1999-04-29
PRIOR APPLICATION NUMBER: 09/188,930
PRIOR FILING DATE: 1998-11-09
PRIOR APPLICATION NUMBER: 09/069,726
PRIOR FILING DATE: 1998-04-29
NUMBER OF SEQ ID NOS: 725
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 438
LENGTH: 1627
TYPE: DNA
ORGANISM: Mouse
US-10-152-661-438

Query Match
Best Local Similarity 51.28; Pred. No. 8.1e-40;
Matches 419; Conservative 0; Mismatches 390; Indels 9; Gaps 2;

Qy 491 CCGGGGCTTGCACATGTCTGCGGGGCTTCCGGGGCTGCGGAGACCGCTGTGAGC 550
Db 1 CCGGAGCTGCTGACTAGCCCTGTGGGTTCTTGGGGCGAGCTGACCTTGTCTCCAC 60
Qy 551 AGGAGCTATGTGTACACATCTCTACAGATGCCAGTCCAGAAATGAGCCACTGCG 610
Db 61 AGGTCGCTTGTGGCCCGCCAGCTGTGCCACGTGTGTATGCGGGGAGGGGCGCATGTG 120
Qy 611 ACCAGCTACGAGGGGAGATGCCCTGCCACAGATACACCGGAGCCCTTGTGTGAGATC 670
Db 121 ACCAGTGTGCGGAGCTGT 180
Qy 671 TTTGTCTCTGTGTAAACATGT 730
Db 181 GCTGTCCCGACGAGCTGT 240

QY 731 GAGTGTGATCATGACGTGAGTGAAGTGTGCTTGGCTTCTGGCTGATGGCAGAGTGT 790
DB 241 GCGTGTGATGCTACCAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300
QY 791 GTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 850
DB 301 GTGAGACAGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360
QY 851 ATATGAGGAGGAGCTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 910
DB 361 AGAACATGAGCAGTGTGAGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
QY 911 GGAAGGAGGAGGAGTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 970
DB 421 GTGACCTGTGAGACAGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480
QY 971 GCGAGTGTGATCAAGGAGGAGTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1030
DB 481 GCGAGTGTGATCAAGGAGGAGTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
QY 1031 GCTTTCCTGCGAGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1090
DB 541 GCTTTCCTGCGAGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 597
QY 1091 GTGACAAAGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1150
DB 598 GCGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 651
QY 1151 AGTGTGCTGCAAGCAGGAGGAGTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1210
DB 652 TGTGCTTGTGCGCAGGAGGAGGAGTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 711
QY 1211 TGTACGGGAGAGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1270
DB 712 GCTTTCCTGCGAGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 771
QY 1271 TGACTGGAAGTGCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1308
DB 772 TCAGTGGGAGTGCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 809

RESULT 15
US-09-866-050A-438
Sequence 438, Application US/09866050A
Publication No. US20030040471A1
GENERAL INFORMATION:
APPLICANT: Watson, James D.
APPLICANT: Strachan, Lorna
APPLICANT: Sleeman, Matthew
APPLICANT: Onrust, Rene
APPLICANT: Murlson, James G.
APPLICANT: Kumble, Krishanand D.
TITLE OF INVENTION: Compositions Isolated From Skin Cells
FILE REFERENCE: 11000.10114U
CURRENT APPLICATION NUMBER: US/09/866,050A
NUMBER OF SEQ ID NOS: 725
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 438
LENGTH: 1627
TYPE: DNA
ORGANISM: Mouse
US-09-866-050A-438

Query Match 4.6%; Score 156; DB 9; Length 1627;
Best Local Similarity 51.2%; Pred. No. 8, 1e-40;
Matches 419; Conservative 0; Mismatches 390; Indels 9; Gaps 2;

QY 491 CCGGAGCCTTGCACAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 550
DB 1 CCGGAGCCTTGCACAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 60

QY 551 AGGACACTATGTGTAACGACTGTGATCAGAGATGCCAGTGTGCTGCTGCTGCTGCTGCTGCT 610
DB 61 AGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120
QY 611 ACCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 670
DB 121 ACCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 180
QY 671 TTTGCTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 730
DB 181 GCTGTCTCCAGAGACCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 240
QY 731 GAGTGTGATCATGACGTGAGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 790
DB 241 GCGTGTGATGATCAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300
QY 791 GTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 850
DB 301 GTGAGACAGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360
QY 851 ATATGAGGAGGAGCTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 910
DB 361 AGAACATGAGCAGTGTGAGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
QY 911 GGAAGGAGGAGGAGTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 970
DB 421 GTGACCTGTGAGACAGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480
QY 971 GCGAGTGTGATCAAGGAGGAGTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1030
DB 481 GCGAGTGTGATCAAGGAGGAGTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
QY 1031 GCTTTCCTGCGAGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1090
DB 541 GCTTTCCTGCGAGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 597
QY 1091 GTGACAAAGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1150
DB 598 GCGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 651
QY 1151 AGTGTGCTGCAAGCAGGAGGAGTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1210
DB 652 TGTGCTTGTGCGCAGGAGGAGGAGTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 711
QY 1211 TGTACGGGAGAGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1270
DB 712 GCTTTCCTGCGAGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 771
QY 1271 TGACTGGAAGTGCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1308
DB 772 TCAGTGGGAGTGCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 809

Search completed: May 9, 2003, 11:53:28
Job time : 467 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4 p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 9, 2003, 06:41:37 ; Search time 118 Seconds
(without alignments)
8896.225 Million cell updates/sec

Title: US-10-092-390-1

Perfect score: 3423

Sequence: 1 atggtattctcttgtaactc.....gcagcagcagcagtgatga 3423

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 1533381 residues

total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents_NA:*

- 1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
- 2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
- 3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
- 4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
- 5: /cgn2_6/ptodata/1/ina/PTUS_COMB.seq:*
- 6: /cgn2_6/ptodata/1/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	length	ID	Description
1	1425.8	41.7	1448	US-09-130-491-9	Sequence 9, Appl
2	165.8	4.8	1578	US-09-130-491-10	Sequence 10, Appl
3	156	4.6	1464	US-09-188-930-255	Sequence 255, Appl
4	155.2	4.5	1633	US-09-188-930-73	Sequence 73, Appl
5	55.2	1.6	4138	US-08-323-474-1	Sequence 1, Appl
6	55.2	1.6	4138	US-08-323-474-1	Sequence 1, Appl
7	54	1.6	3845	US-08-220-240A-4	Sequence 1, Appl
8	54	1.6	4175	US-07-934-393B-1	Sequence 4, Appl
9	54	1.6	4175	US-08-278-089A-1	Sequence 1, Appl
10	54	1.6	4175	US-08-278-089A-1	Sequence 1, Appl
11	54	1.6	4176	US-08-378-089A-5	Sequence 5, Appl
12	54	1.6	4176	US-08-378-089A-5	Sequence 5, Appl
13	49	1.4	3489	US-08-728-323A-1	Sequence 1, Appl
14	49	1.4	3489	US-08-728-323A-1	Sequence 1, Appl
15	49	1.4	32207	US-08-770-379-20	Sequence 20, Appl
16	49	1.4	32207	US-08-770-379-20	Sequence 20, Appl
17	49	1.4	32207	US-08-757-669A-20	Sequence 20, Appl
18	44.4	1.3	4411529	US-09-230-371A-20	Sequence 20, Appl
19	40.6	1.2	2144	US-09-103-840A-1	Sequence 1, Appl
20	40.6	1.2	2144	US-08-834-306-15	Sequence 15, Appl
21	40.6	1.2	2144	US-09-256-976-15	Sequence 15, Appl
22	39.4	1.2	477	US-09-135-994-1	Sequence 1, Appl
23	39	1.1	1235	US-08-557-309B-53	Sequence 53, Appl
24	37.4	1.1	2640	US-08-684-932A-37	Sequence 37, Appl
25	36.2	1.1	1876	US-08-483-232-24	Sequence 24, Appl
26	36.2	1.1	1876	US-08-485-938A-24	Sequence 24, Appl
27	36.2	1.1	1876	US-08-910-041-24	Sequence 24, Appl

c 28	36.2	1.1	1876	US-09-328-474-24	Sequence 24, Appl
c 29	36.2	1.1	1876	US-09-100-546-24	Sequence 24, Appl
c 30	36.2	1.1	1876	US-09-010-715-24	Sequence 24, Appl
c 31	36.2	1.1	1876	US-09-577-758-24	Sequence 24, Appl
c 32	35.8	1.0	1839	US-08-122-458B-10	Sequence 10, Appl
c 33	35.8	1.0	6294	US-08-976-259-5	Sequence 5, Appl
c 34	35.4	1.0	2625	US-09-245-041-18	Sequence 18, Appl
c 35	35.4	1.0	4072	US-09-245-041-16	Sequence 16, Appl
c 36	35.4	1.0	5118	US-08-669-785-3	Sequence 3, Appl
c 37	35.4	1.0	8257	US-09-484-970B-65	Sequence 65, Appl
c 38	35.4	1.0	8589	US-09-245-041-14	Sequence 14, Appl
c 39	35	1.0	290	US-08-818-112-37	Sequence 37, Appl
c 40	35	1.0	290	US-08-818-111-37	Sequence 37, Appl
c 41	35	1.0	290	US-09-056-556-37	Sequence 37, Appl
c 42	35	1.0	290	US-09-072-596-37	Sequence 37, Appl
c 43	35	1.0	1743	US-08-665-259-20	Sequence 20, Appl
c 44	35	1.0	1743	US-08-762-500-20	Sequence 20, Appl
c 45	35	1.0	1974	US-08-762-500-78	Sequence 78, Appl

ALIGNMENTS

RESULT 1	US-09-130-491-9	Sequence 9, Application US/09130491
Patent No. 6416974		
GENERAL INFORMATION:		
APPLICANT: Holtzman, Douglas A.		
APPLICANT: Goodearl, Andrew D.J.		
TITLE OF INVENTION: TANGO-71, TANGO-73, TANGO-74, TANGO-76, AND TANGO-83		
FILE REFERENCE: 09404/041001		
CURRENT APPLICATION NUMBER: US/09/130,491		
EARLIER FILING DATE: 1998-08-07		
EARLIER APPLICATION NUMBER: US 60/056,108		
EARLIER FILING DATE: 1997-09-05		
EARLIER APPLICATION NUMBER: US 60/054,961		
NUMBER OF SEQ ID NOS: 16		
SOFTWARE: FASTSEQ for Windows Version 3.0		
SEQ ID NO 9		
LENGTH: 1448		
TYPE: DNA		
ORGANISM: Homo sapiens		
US-09-130-491-9		
Query Match	41.7%	Score 1425.8; DB 4; Length 1448;
Best Local Similarity	99.9%	Pred. No. 0;
Matches 1427; Conservative	0;	Mismatches 2; Indels 0; Gaps 0;
QY 1216	GGGGAACCTTGGCAGACGATCTGCGCAAAATGGGCGACACGTGTGACGTGACT	77
DB 18	GGGGAACCTTGGCAGACGATCTGCGCAAAATGGGCGACACGTGTGACGTGACT	77
QY 1276	GGAAAGTGCACCTGTGCGCCAGATTCAAAGATTCAGTCTTACCCCATGCTCTG	1335
DB 78	GGAAAGTGCACCTGTGCGCCAGATTCAAAGATTCAGTCTTACCCCATGCTCTG	137
QY 1336	GGAACCTATGGATTAACCTTCTCGCTGTGGCTGTAAATGATGACGTCTCTCT	1395
DB 138	GGAACCTATGGATTAACCTTCTCGCTGTGGCTGTAAATGATGACGTCTCTCT	197
QY 1396	CCGTGTGACGCGTCTTGTACTTGCAGAGCGCTGCGACGCGGTGACACTCTCATCAGA	1455
DB 198	CCGTGTGACGCGTCTTGTACTTGCAGAGCGCTGCGACGCGGTGACACTCTCATCAGA	257
QY 1456	TGTCCAGTGGCAGATGGGCTTGGCTGTAACTTAACATGCCAGTGCCTCAACGGGGA	1515
DB 258	TGTCCAGTGGCAGATGGGCTTGGCTGTAACTTAACATGCCAGTGCCTCAACGGGGA	317
QY 1516	GCCTGCAACACCTGGAACGGGACCTGCACTGTGACACCTGATGGCGGGGGAATGC	1575
DB 318	GCCTGCAACACCTGGAACGGGACCTGCACTGTGACACCTGATGGCGGGGGAATGC	377

Best Local Similarity 51.2%; Pred. No. 9.7e-38;
Matches 419; Conservative 0; Mismatches 390; Indels 9; Gaps 2;

```
Oy 491 CCGGGGCTTGGCAGTGTCTGCGGGCTTCGGGGCTGGCGAGACCGCTGTAGC 550
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9 CCGGAGCTGTGTAAGTCTGCTGGTTCCTTGGGGCCGAGTGTAGCCTTGTGTCCAC 68
Oy 551 AGGGACCTATGATGATGACAGTGTCTGAGATGCGCATGCGAGATGAGACCGACTGCG 610
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 69 AGGGCTGTGGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 128
Oy 611 ACCAGTCAGGGGGGAGATGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 670
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 129 ACCAGTCAGGGGGGAGTGTCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 188
Oy 671 TTTGTCTCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 730
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 189 GCTGTCTCTGAGGAGCGGTTTGGCAAGGCGTGTGAACAGATGTGCTGTGAGGATGAGG 248
Oy 721 GAGTGTCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 790
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 249 GCTGTGTCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 308
Oy 791 GTGTGACGCTTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 850
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 309 GTGAGACGCGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 368
Oy 851 ATATGAGGAGGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 910
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 369 AGAACATGAGGAGGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 428
Oy 911 GGAAGAGGTCGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 970
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 429 GTCAAGCTTGTGAGAGACACCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 488
Oy 971 GCCAGTGTGTCAGGAGGAGGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATG 1030
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 489 GCGAGTGTCAACAGGAGGAGGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATG 548
Oy 1031 GCTTGTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1090
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 549 GCTTGTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 605
Oy 1091 GTGACAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1150
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 606 GCTGTGACAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 659
Oy 1151 AGTGTGCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1210
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 660 TCTGTGCTTGTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 719
Oy 1211 TCTAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1270
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 720 GCTTGTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 779
Oy 1271 TGACTGAAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1308
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 780 TCAGTGGGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 817
```

RESULT 4

US-09-188-930-73

: Sequence 73, Application US/09188930A

: Patent No. 6150502

: GENERAL INFORMATION:

: APPLICANT: Watson, James D.

: APPLICANT: Strachan, Lorna

: APPLICANT: Sleeman, Matthew

: APPLICANT: Murison, James Greg

: TITLE OF INVENTION: Compositions Isolated From Skin Cells

: FILE REFERENCE: 11000.1011c1

: CURRENT APPLICATION NUMBER: US/09/188,930A
: CURRENT FILING DATE: 1998-11-09
: NUMBER OF SEQ ID NOS: 348
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO: 73

: TYPE: DNA
: ORGANISM: mouse
: FEATURE:
: NAME/KEY: unsure
: LOCATION: (1608)...(1608)
US-09-188-930-73

Query Match 4.5%; Score 155.2; DB 3; Length 1633;

Best Local Similarity 51.0%; Pred. No. 1.8e-37; Matches 417; Conservative 2; Mismatches 390; Indels 9; Gaps 2;

```
Oy 491 CCGGGGCTTGGCAGTGTCTGCGGGCTTCGGGGCTGGCGAGACCGCTGTAGC 550
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9 CCGGAGCTGTGTAAGTCTGCTGGTTCCTTGGGGCCGAGTGTAGCCTTGTGTCCAC 68
Oy 551 AGGGACCTATGATGATGACAGTGTCTGAGATGCGCATGCGAGATGAGACCGACTGCG 610
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 69 AGGGCTGTGGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 128
Oy 611 ACCAGTCAGGGGGGAGATGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 670
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 129 ACCAGTCAGGGGGGAGTGTCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 188
Oy 671 TTTGTCTCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 730
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 189 GCTGTCTCTGAGGAGCGGTTTGGCAAGGCGTGTGAACAGATGTGCTGTGAGGATGAGG 248
Oy 721 GAGTGTCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 790
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 189 GCTGTCTCTGAGGAGCGGTTTGGCAAGGCGTGTGAACAGATGTGCTGTGAGGATGAGG 248
Oy 731 GAGTGTCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 790
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 249 GCTGTGTCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 308
Oy 791 GTGTGACGCTTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 850
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 309 GTGAGACGCGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 368
Oy 851 ATATGAGGAGGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 910
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 369 AGAACATGAGGAGGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 428
Oy 911 GGAAGAGTGTGTCAGGAGGAGGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATG 1030
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 429 GTCAAGCTTGTGAGAGACACCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 488
Oy 971 GCCAGTGTGTCAGGAGGAGGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1030
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 489 GCGAGTGTCAACAGGAGGAGGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATG 548
Oy 1031 GCTTGTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1090
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 549 GCTTGTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 605
Oy 1091 GTGACAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1150
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 606 GCTGTGACAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 659
Oy 1151 AGTGTGCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1210
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 660 TCTGTGCTTGTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 719
Oy 1211 TCTAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1270
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 720 GCTTGTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 779
Oy 1271 TGACTGAAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1308
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 780 TCAGTGGGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 817
```

```

ADDRESS: Immunex Corporation
STREET: 51 University Street
City: Seattle
STATE: Washington
COUNTRY: US
ZIP: 98101

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/06093
FILING DATE: 19930625
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/905,600
FILING DATE: 26-JUN-1992
ATTORNEY/AGENT INFORMATION:
NAME: Seese, Kathryn A.
REGISTRATION NUMBER: 32,172
REFERENCE/DOCKET NUMBER: 2609
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 587-0430
TELEFAX: (206) 233-0644
TELEX: 756822

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4138 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
FEATURE:
NAME/KEY: CDS
LOCATION: 149..3523
PCT-US93-06093-1

Query Match 1.6%, Score 55.2, DB 5, Length 4138;
Best Local Similarity 60.8%; Pred. No. 2.8e-06;
Matches 90; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 1131 CTGTACCCCATGTGTGGAGATGTGCTGCAGAGCGGGCTGTGCAGACCTTACTGTAA 1190
DB 985 CTGTCTCCCTGACACCCCTATGGGTGTCTCTGTGCCACAGGCTGGAAAGGCTTCACAGTGCA 1044
QY 1191 TGAGACATGTTCTCTCGATTTCTACGGGGAAAGCTTGCCAGACGATTCGACGTGCCAAA 1250
DB 1045 TGAAGCATGCGCCACCTGGTGTTTTACGGGCGACAGATTGTAAAGTTAGTGTGACGTGCACAA 1104
QY 1251 TGGGGACAGCTGTGTGACAGTGTGACTGGA 1278
DB 1105 TGGGAGATGTGTGATCGCTTCCAGGA 1132

RESULT 7
US-08-220-240A-4
Sequence 4, Application US/08220240A
Patent No. 5955291
GENERAL INFORMATION:
APPLICANT: Aitalo, Karl
APPLICANT: Matikainen, Marja-Terttu
APPLICANT: Partanen, Juha
APPLICANT: Makiela, Tomi
APPLICANT: Korhonen, Jaana
TITLE OF INVENTION: ANTIBODIES RECOGNIZING THE RECEPTOR
TITLE OF INVENTION: TYROSINE KINASE AND USES THEREOF
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 233 South Wacker Drive/6300 Sears Tower
City: Chicago
STATE: Illinois

```


Db 920 AAGTGTAGTGGACAGAGATGCAAGCTTATGTGTCTCCCAAGCCTTACG 979
 QY 884 AATGTCATTCAGATCCAGATACACAGGAGGATGCCAGATGATGCTGTGGGA 943
 Db 980 GGTGTCCTGTCGACAGCTGAGAGGGGTGTCAGTCAATGAAAGATCCCATCGGT 1039
 QY 944 CCTATGCGCTTCTGTGTGTGAGACCTGCCAGTGTCTCAAGGAGGAAGTTTACG 1003
 Db 1040 ACTAGCGACACAGCTGTAGCTCAGGTGCCACTGTACCAATGAAAGATATGATCGGT 1099
 QY 1004 TGAGCGGCGCATGCTCTGTGAAGCAG 1031
 Db 1100 TCCAGGATGCTCTGCTCTCAAGATG 1127

RESULT 9
 US-08-278-089A-1
 ; Sequence 1, Application US/08278089A
 ; Patent No. 5681714
 ; GENERAL INFORMATION:
 ; APPLICANT: Breitman, Martin L.
 ; APPLICANT: Rossant, Janet
 ; APPLICANT: Dumont, Daniel J.
 ; APPLICANT: Yamaguchi, Terry P.
 ; TITLE OF INVENTION: No. 5681714e1 Receptor Tyrosine Kinase
 ; NUMBER OF SEQUENCES: 33
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Bereskin & Parr
 ; STREET: 40 King Street West
 ; CITY: Toronto
 ; STATE: Ontario
 ; COUNTRY: Canada
 ; ZIP: M5H 3Y2
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; OPERATING SYSTEM: IBM PC compatible
 ; SOFTWARE: Patentin Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/278,089A
 ; FILING DATE: 20-JUL-1994
 ; CLASSIFICATION: 530
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Kurdydky, Linda M.
 ; REGISTRATION NUMBER: 34,971
 ; REFERENCE/DOCKET NUMBER: 3153-111
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (416) 364-7311
 ; TELEFAX: (416) 361-1398
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 4175 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: CDNA
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: NO
 ; FRAGMENT TYPE: N-terminal
 ; ORIGINAL SOURCE:
 ; ORGANISM: Mus musculus.
 ; STRAIN: CD-1
 ; DEVELOPMENTAL STAGE: Embryo
 ; TISSUE TYPE: Heart
 ; IMMEDIATE SOURCE:
 ; CLONE: Tek
 ; POSITION IN GENOME:
 ; CHROMOSOME/SEGMENT: 4
 ; MAP POSITION: Between the brown and pmv-23 loci
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 124..3478
 ; US-08-278-089A-1

Query Match 1.6%; Score 54; DB 1; Length 4175;
 Best Local Similarity 50.6%; Pred. No. 6.5e-06;
 Matches 166; Conservative 0; Mismatches 150; Indels 12; Gaps 1;

QY 716 CTTTCAGAAATGAGAGAGTGTGTCATCACGTCACGTGAGAAATGCTTCCCTTGGCT 775
 Db 800 CTTCAGAAACAAATGAGAGTGTGTCATGAAAGCTTGTGAGCCGACACATTTGGCAGACCTTAAG 919
 QY 776 GGATGGGACAGTGTGTGTCAGCTTGCCCGGAGGGTGGCTTGGAAAGAACTGTGCC 835
 Db 860 TTATGGGAGAGACATGTGAGAAAGCTTGTGAGCCGACACATTTGGCAGACCTTAAG 919
 QY 836 AAGATGCCAGTGCCTAATATGAGGAGCGTGTGATG-----CTGCCACAGGCC 883
 Db 920 AAGGTGTAGTGACACAGAGATGCAAGTCTTATGTGTCTCCCAAGACCTTACG 979
 QY 884 AATGTCATTCAGATCCAGATACACAGGAGGAAAGGTCGCAAGATAGTCTGTGGGA 943
 Db 980 GGTGTCCTGTCGACAGCTGAGAGGGGTTCAGTGCATGCAATGAAAGATGCCCATCGGT 1039
 QY 944 CCTATGCGCTTCTGTGTGTGAGACCTGCCAGTGTCTCAAGGAGGAAGTTTACG 1003
 Db 1040 ACTAGCGACACAGCTGTAGCTCAGGTGCCACTGTACCAATGAAAGATATGATCGGT 1099
 QY 1004 TGAGCGGCGCATGCTCTGTGAAGCAG 1031
 Db 1100 TCCAGGATGCTCTGCTCTCAAGATG 1127

RESULT 10
 US-08-838-957A-1
 ; Sequence 1, Application US/08838957A
 ; Patent No. 598187
 ; GENERAL INFORMATION:
 ; APPLICANT: Breitman, Martin L.
 ; APPLICANT: Rossant, Janet
 ; APPLICANT: Dumont, Daniel J.
 ; APPLICANT: Yamaguchi, Terry P.
 ; TITLE OF INVENTION: No. 598187e1 Receptor Tyrosine Kinase
 ; NUMBER OF SEQUENCES: 32
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Bereskin & Parr
 ; STREET: 40 King Street West
 ; CITY: Toronto
 ; STATE: Ontario
 ; COUNTRY: Canada
 ; ZIP: M5H 3Y2
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; OPERATING SYSTEM: IBM PC compatible
 ; SOFTWARE: Patentin Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/838,957A
 ; FILING DATE: 23-APR-1997
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Kurdydky, Linda M.
 ; REGISTRATION NUMBER: 34,971
 ; REFERENCE/DOCKET NUMBER: 3153-212
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (416) 364-7311
 ; TELEFAX: (416) 361-1398
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 4175 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: CDNA
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: NO


```

STREET: 40 King Street West
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5H 3Y2
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/838,957A
FILING DATE: 23-APR-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Kurdzyk, Linda M.
REGISTRATION NUMBER: 34,971
REFERENCE/DOCKET NUMBER: 3153-212
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 364-7311
TELEFAX: (416) 361-1398
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 4176 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
AMT-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: Mus musculus
STRAIN: CD-1
DEVELOPMENTAL STAGE: Embryo
TISSUE TYPE: Heart
IMMEDIATE SOURCE:
CLONE: Tek
POSITION IN GENOME:
CHROMOSOME/SEGMENT: 4
MAP POSITION: Between the brown and pmv-23 loci
FEATURE:
NAME/KEY: CDS
LOCATION: 124..3490
US-08-838-957A-5

Query Match
Best Local Similarity 1.6%; Score 54; DB 2; Length 4176;
Matches 166; Conservative 0; Mismatches 150; Indels 12; Gaps 1;

```

Query 716 CTTGTCAAAATGAGAGAGTGTGTATCAGTCACTGAGAGATGCTTGCCTTCGTGCT 775
 DB 800 CTTGCAAGAACATGAGTGTCCATGAAGATACCGGGGATGCAATTCCTCCGTGGGT 859
 QY 776 GGATGGGACATGTGTGTGACCTTGCCCGAGGGTGGCTTTGGAAGAACTGTCCC 835
 DB 860 TTATGGGGGAACATGTGGAAGCTTGTGACCCGACACATTTGGCAGACCTGTAAAG 919
 QY 836 AAGATGCGCAGTGCATATATGAGGAGAGCTGTGATG-----CTGCCACAGGCC 883
 DB 920 AAGGTGTAGTGGACACAGAGATGCAACTTATGTGTCTGTCTCCAGACCTTACG 979
 QY 884 AATGTCATTCAGTGCAGATACACAGGGAGGATGCCAGATGAGTCTCTGTGGGA 943
 DB 980 GGTGTCTGTGTCACAGAGCTGGAGGGGTGCAGTGCATGAAGATGCAATCCATCGT 1039
 QY 944 CTTATGGCGTCTGTGTGTGAGACCTGCCAGTGTGTCAACGAGGAGAGTTTACACAG 1003
 DB 1040 ACTAGGACACAGCTGTAGCTCAGGTGCTACCAATGAAGATATGTGATCGGT 1099
 QY 1004 TGAGCGGCGCATGCTGTGAAGCAGG 1031
 DB 1100 TCCAAGGATGCTGTCTCAAGCATG 1127

```

RESULT 13
US-08-728-323A-1/c
Sequence 1, Application US/08728323A
Patent No. 5948676
GENERAL INFORMATION:
APPLICANT: Chang, Yuan
APPLICANT: Bohenzky, Roy A.
APPLICANT: Russo, James J.
APPLICANT: Edelman, Isidore S.
APPLICANT: Moore, Patrick S.
TITLE OF INVENTION: Immediate Early Protein From Kaposi's
TITLE OF INVENTION: Sarcoma-Associated Herpesvirus, DNA
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/728,323A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 0575/52268/JPM/MSC/SKS
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-278-0400
TELEFAX: 212-391-0525
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3489 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..3489
US-08-728-323A-1

Query Match
Best Local Similarity 1.4%; Score 49; DB 2; Length 3489;
Matches 151; Conservative 0; Mismatches 170; Indels 0; Gaps 0;

```

Query 1622 GCTGCGACTGACGACCAAGAGTGTGACACCTTACACAGGGGCAATGCGCGCTCC 1681
 DB 1726 GCTGTGGCTCCCGCTGTGTGCTCCCGCTGTGGCTCCCGCTGTGGGCTCCCGCT 1667
 QY 1682 CGGATGTGACAGTGTCCACTGTGACAGCTGTGTGTGTGAGGAGAGCGGGGCCCACT 1741
 DB 1666 GCTGTGGCTCCCGCTGTGTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1607
 QY 1742 GCTCCCTCCCTGACTGTAATAAATGAGGCTTCAATGCTCCCGTATGATGCACTGCG 1801
 DB 1606 GCTGTGGCTCCCGCTGT 1547
 QY 1802 AGTGTGACCAAGCTTCCGAGGACCACTTGTCAAGAGATGCTCCCGGTTTATG 1861
 DB 1546 GCTGTGGCTCCCGCTGT 1487
 QY 1862 GGCATGCTGACGACCAAGATGCCCAAGTGTTCACAGCAGCGGGGCCCTGCCACCA 1921

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4.P5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 9, 2003, 10:55:56 ; Search time 18.4936 Seconds

(Without alignments)
2556.719 Million cell updates/sec

Title: US-10-092-390-2
Sequence: 1 MVSLNSCLSFICLLCHWI.....SSPKQEDSGSSSSSSSSSE 1140

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1037	15.4	2524	1	NOTC_XENLA
2	1034.5	15.3	2536	1	NOTC1_HUMAN
3	1028	15.2	2531	1	NOTC1_MOUSE
4	1024	15.2	2531	1	NOTC1_RAT
5	1014.5	15.0	2471	1	NOTC2_HUMAN
6	998	14.8	2471	1	NOTC2_MOUSE
7	993	14.7	2470	1	NOTC2_MOUSE
8	987	14.6	2437	1	NOTC1_BRARE
9	978.5	14.5	2703	1	NOTC1_MOUSE
10	977.5	14.5	2318	1	NOTC3_MOUSE
11	974	14.4	2321	1	NOTC3_HUMAN
12	969.5	14.2	2319	1	NOTC3_RAT
13	959.5	14.2	2003	1	NOTC4_HUMAN
14	954.5	14.2	1064	1	FBP1_STRPU
15	951.5	14.1	1964	1	NOTC4_MOUSE
16	916.5	13.6	4289	1	TENX_HUMAN
17	870.5	12.9	830	1	SREC_HUMAN
18	832.5	12.3	1213	1	JAG3_BRARE
19	789	11.7	1238	1	JAG2_HUMAN
20	775.5	11.5	2139	1	CRB_DROME
21	775	11.5	1242	1	JAG1_BRARE
22	769	11.4	1201	1	TENA_HUMAN
23	768	11.4	1247	1	JAG2_MOUSE
24	767.5	11.4	1246	1	TENA_PIG
25	757	11.2	1218	1	JAG1_HUMAN
26	745	11.0	1202	1	JAG2_RAT
27	744	11.0	1218	1	JAG1_RAT
28	739	11.0	1218	1	JAG1_MOUSE
29	736.5	10.9	3695	1	LMAS_HUMAN
30	720.5	10.7	1808	1	TENA_CHICK
31	717.5	10.6	1801	1	LMB2_RAT
32	716	10.6	3106	1	LMB2_MOUSE
33	706.5	10.5	1790	1	LMB1_DROME

ALIGNMENTS

34	704.5	10.4	3084	1	LM1_MOUSE
35	700.5	10.4	1799	1	LM2_MOUSE
36	692	10.3	3075	1	LM1_HUMAN
37	690	10.2	1408	1	SERR_DROME
38	687.5	10.2	3110	1	LM2_HUMAN
39	685.5	10.2	1798	1	LM2_HUMAN
40	683.5	10.1	3718	1	LMAS_MOUSE
41	676.5	10.0	3672	1	LM1_MOUSE
42	667	9.9	1786	1	LM1_HUMAN
43	664	9.8	1786	1	LM1_MOUSE
44	657	9.7	1429	1	LM1_MOUSE
45	638	9.5	3712	1	LM1_DROME

RESULT 1
NOTC_XENLA
ID NOTC_XENLA STANDARD: PRT; 2524 AA.

AC P21783;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Neurogenic locus notch protein homolog precursor (XORCH protein).
GN XORCH
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
OC Xenopodidae; Xenopus.
OX NCBI_Taxid=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90385285; PubMed=2402639;
RA Ciftman C., Harris W., Kintner C.;
RT "Xorch, the Xenopus homolog of Drosophila notch.";
RL Science 249:1438-1441(1990).
RN [2]
RP REVISIONS TO 1759-1782.

RA Kintner C.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED ALMOST UNIFORMLY IN EARLY EMBRYOS.
CC -1- SIMILARITY: BELONGS TO THE NOTCH FAMILY.
CC -1- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC -1- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M33874; AB02039.1; -
DR PIR: A35844; A35844.
DR HSP: P00740; IEDM.
DR InterPro: IPR002110; ANK.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000561; TENA-like.
DR InterPro: IPR000742; EGF_2.
DR InterPro: IPR001881; EGF_CA.
DR InterPro: IPR001438; EGF_II.
DR InterPro: IPR000800; Notch.
DR Pfam: PF000023; EGF; 36.
DR Pfam: PF000066; notch; 3.
DR PRINTS: PR00010; EGFBLD.
DR PRINTS: PR01452; NOTCH.
DR SMART: SM00248; ANK; 5.
DR SMART: SM00179; EGF_CA; 23.

FT	DISULFID	188	203	BY SIMILARITY
FT	DISULFID	205	214	BY SIMILARITY
FT	DISULFID	221	232	BY SIMILARITY
FT	DISULFID	226	242	BY SIMILARITY
FT	DISULFID	244	253	BY SIMILARITY
FT	DISULFID	260	271	BY SIMILARITY
FT	DISULFID	265	280	BY SIMILARITY
FT	DISULFID	282	291	BY SIMILARITY
FT	DISULFID	298	311	BY SIMILARITY
FT	DISULFID	305	320	BY SIMILARITY
FT	DISULFID	322	331	BY SIMILARITY
FT	DISULFID	328	349	BY SIMILARITY
FT	DISULFID	422	437	BY SIMILARITY
FT	DISULFID	439	448	BY SIMILARITY
FT	DISULFID	455	466	BY SIMILARITY
FT	DISULFID	460	475	BY SIMILARITY
FT	DISULFID	477	486	BY SIMILARITY
FT	DISULFID	493	504	BY SIMILARITY
FT	DISULFID	498	513	BY SIMILARITY
FT	DISULFID	515	524	BY SIMILARITY
FT	DISULFID	531	542	BY SIMILARITY
FT	DISULFID	536	551	BY SIMILARITY
FT	DISULFID	553	562	BY SIMILARITY
FT	DISULFID	569	579	BY SIMILARITY
FT	DISULFID	574	588	BY SIMILARITY
FT	DISULFID	590	599	BY SIMILARITY
FT	DISULFID	606	617	BY SIMILARITY
FT	DISULFID	611	626	BY SIMILARITY
FT	DISULFID	628	637	BY SIMILARITY
FT	DISULFID	644	654	BY SIMILARITY
FT	DISULFID	649	663	BY SIMILARITY
FT	DISULFID	665	674	BY SIMILARITY
FT	DISULFID	681	692	BY SIMILARITY
FT	DISULFID	686	701	BY SIMILARITY
FT	DISULFID	703	712	BY SIMILARITY
FT	DISULFID	719	729	BY SIMILARITY
FT	DISULFID	724	738	BY SIMILARITY
FT	DISULFID	740	749	BY SIMILARITY
FT	DISULFID	756	767	BY SIMILARITY
FT	DISULFID	761	776	BY SIMILARITY
FT	DISULFID	778	787	BY SIMILARITY
FT	DISULFID	794	805	BY SIMILARITY
FT	DISULFID	799	814	BY SIMILARITY
FT	DISULFID	816	825	BY SIMILARITY
FT	DISULFID	832	843	BY SIMILARITY
FT	DISULFID	837	854	BY SIMILARITY
FT	DISULFID	856	863	BY SIMILARITY
FT	DISULFID	872	883	BY SIMILARITY
FT	DISULFID	877	892	BY SIMILARITY
FT	DISULFID	884	903	BY SIMILARITY
FT	DISULFID	910	921	BY SIMILARITY
FT	DISULFID	915	930	BY SIMILARITY
FT	DISULFID	932	941	BY SIMILARITY
FT	DISULFID	966	997	BY SIMILARITY
FT	DISULFID	991	1007	BY SIMILARITY
FT	DISULFID	1008	1017	BY SIMILARITY
FT	DISULFID	1024	1035	BY SIMILARITY
FT	DISULFID	1029	1044	BY SIMILARITY
FT	DISULFID	1046	1055	BY SIMILARITY
FT	DISULFID	1062	1073	BY SIMILARITY
FT	DISULFID	1067	1082	BY SIMILARITY
FT	DISULFID	1084	1093	BY SIMILARITY
FT	DISULFID	1100	1121	BY SIMILARITY
FT	DISULFID	1115	1131	BY SIMILARITY
FT	DISULFID	1132	1141	BY SIMILARITY
FT	DISULFID	1148	1159	BY SIMILARITY
FT	DISULFID	1153	1168	BY SIMILARITY

CC		bra stem and lung. Also present in most adult tissues where it
CC		is found mainly in lymphoid tissues.
CC	-1-	PTM: Synthesized in the endoplasmic reticulum as an inactive form
CC		which is proteolytically cleaved by a furin-like convertase in the
CC		trans-golgi network before it reaches the plasma membrane to yield
CC		an active, ligand-accessible form. Cleavage results in a C-
CC		terminal fragment N(TM) and a N-terminal fragment N(EC). Following
CC		ligand binding, it is cleaved by TNF-alpha converting enzyme
CC		(TACE) to yield a membrane-associated intermediate fragment called
CC		notch extracellular truncation (NEXT). This fragment is then
CC		cleaved by presenilin dependent gamma-secretase to release a
CC		notch-derived peptide containing the intracellular domain (NICD)
CC		from the membrane (By similarity).
CC	-1-	PTM: Phosphorylated (By similarity).
CC	-1-	DISEASE: Notch1 truncation is associated with neoplasia, a T-cell
CC		acute lymphoblastic leukemia.
CC	-1-	SIMILARITY: BELONGS TO THE NOTCH FAMILY.
CC	-1-	SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC	-1-	SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC	-1-	SIMILARITY: CONTAINS 5 ANK REPEATS.
CC		-----
CC		This SWISS-PROT entry is copyright. It is produced through a collaboration
CC		between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC		the European Bioinformatics Institute. There are no restrictions on its
CC		use by non-profit institutions as long as its content is in no way
CC		modified and this statement is not removed. Usage by and for commercial
CC		entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC		or send an email to license@lsb.sib.ch).
CC		-----
DR	EMBL:	AF308602; AAC3848.1; -
DR	EMBL:	M73980; AAA60614.1; -
DR	HSSP:	P00740; IEDM.
DR	Genew:	HCNC:7881; NOTCH1.
DR	MIT:	190198; --
DR	InterPro:	IPR002110; ANK.
DR	InterPro:	IPR000152; Asx_hydroxy1.
DR	InterPro:	IPR000561; EGF_1like.
DR	InterPro:	IPR000742; EGF_2.
DR	InterPro:	IPR001881; EGF_Ca.
DR	InterPro:	IPR001438; EGF_11.
DR	InterPro:	IPR000800; Notch.
DR	Pfam:	PF00008; EGF; 36.
DR	Pfam:	PF00023; ank; 6.
DR	Pfam:	PF00066; notch; 3.
DR	SMART:	SMO0248; ANK; 5.
DR	SMART:	SMO0179; EGF_CA; 22.
DR	SMART:	SMO0001; EGF_1like; 13.
DR	SMART:	SMO0004; NL; 2.
DR	PROSITE:	PS50088; ANK_REPEAT; 4.
DR	PROSITE:	PS50297; ANK_REP_REGION; 1.
DR	PROSITE:	PS00010; ASX_HYDROXYL; 20.
DR	PROSITE:	PS00022; EGF_1; 34.
DR	PROSITE:	PS01186; EGF_2; 26.
DR	PROSITE:	PS01187; EGF_CA; 18.
KW	Receptor:	Transcription regulation; Activator; Differentiation;
KW	Developmental protein:	Repeat; ANK repeat; EGF-like domain;
KW	Transmembrane:	Glycoprotein; Signal; Phosphorylation.
FT	SIGNAL	1
FT	CHAIN	19 2556
FT	CHAIN	1722 2556
FT	CHAIN	1755 2556
FT	DOMAIN	19 1736
FT	TRANSMEM	1737 1757
FT	DOMAIN	1758 2556
FT	DOMAIN	20 58
FT	DOMAIN	59 99
FT	DOMAIN	102 139
FT	DOMAIN	140 176
FT	DOMAIN	178 216
FT	DOMAIN	218 255
FT	DOMAIN	257 293

FT	DOMAIN	295	333	EGF-LIKE 8,	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	335	371	EGF-LIKE 9,	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	372	410	EGF-LIKE 10,	EGF-LIKE 10.
FT	DOMAIN	412	450	EGF-LIKE 11,	CALCIUM-BINDING (POTENTIAL)
FT	DOMAIN	452	488	EGF-LIKE 12,	CALCIUM-BINDING (POTENTIAL)
FT	DOMAIN	490	526	EGF-LIKE 13,	CALCIUM-BINDING (POTENTIAL)
FT	DOMAIN	528	566	EGF-LIKE 14,	CALCIUM-BINDING (POTENTIAL)
FT	DOMAIN	566	601	EGF-LIKE 15,	CALCIUM-BINDING (POTENTIAL)
FT	DOMAIN	603	639	EGF-LIKE 16,	CALCIUM-BINDING (POTENTIAL)
FT	DOMAIN	641	676	EGF-LIKE 17,	CALCIUM-BINDING (POTENTIAL)
FT	DOMAIN	678	714	EGF-LIKE 18,	CALCIUM-BINDING (POTENTIAL)
FT	DOMAIN	716	751	EGF-LIKE 19,	CALCIUM-BINDING (POTENTIAL)
FT	DOMAIN	753	789	EGF-LIKE 20,	EGF-LIKE 20.
FT	DOMAIN	791	827	EGF-LIKE 21,	CALCIUM-BINDING (POTENTIAL)
FT	DOMAIN	829	868	EGF-LIKE 22,	EGF-LIKE 22.
FT	DOMAIN	870	906	EGF-LIKE 23,	CALCIUM-BINDING (POTENTIAL)
FT	DOMAIN	908	944	EGF-LIKE 24,	EGF-LIKE 24.
FT	DOMAIN	946	982	EGF-LIKE 25,	CALCIUM-BINDING (POTENTIAL)
FT	DOMAIN	984	1020	EGF-LIKE 26,	EGF-LIKE 26.
FT	DOMAIN	1022	1058	EGF-LIKE 27,	EGF-LIKE 27.
FT	DOMAIN	1060	1096	EGF-LIKE 28,	EGF-LIKE 28.
FT	DOMAIN	1098	1144	EGF-LIKE 29,	EGF-LIKE 29.
FT	DOMAIN	1146	1182	EGF-LIKE 30,	EGF-LIKE 30.
FT	DOMAIN	1184	1220	EGF-LIKE 31,	CALCIUM-BINDING (POTENTIAL)
FT	DOMAIN	1222	1266	EGF-LIKE 32,	CALCIUM-BINDING (POTENTIAL)
FT	DOMAIN	1268	1306	EGF-LIKE 33,	EGF-LIKE 33.
FT	DOMAIN	1308	1347	EGF-LIKE 34,	EGF-LIKE 34.
FT	DOMAIN	1349	1385	EGF-LIKE 35,	EGF-LIKE 35.
FT	DOMAIN	1388	1427	EGF-LIKE 36,	EGF-LIKE 36.
FT	REPEAT	1446	1481	LINE/NOTCH 1,	LINE/NOTCH 1.
FT	REPEAT	1482	1523	LINE/NOTCH 2,	LINE/NOTCH 2.
FT	REPEAT	1524	1563	LINE/NOTCH 3,	LINE/NOTCH 3.
FT	REPEAT	1928	1957	ANK 1,	ANK 1.
FT	REPEAT	1961	1991	ANK 2,	ANK 2.
FT	REPEAT	1995	2024	ANK 3,	ANK 3.
FT	REPEAT	2028	2057	ANK 4,	ANK 4.
FT	REPEAT	2061	2090	ANK 5,	ANK 5.
FT	DOMAIN	1576	1579	POLY-VAL,	POLY-VAL.
FT	DOMAIN	1662	1665	POLY-ARG,	POLY-ARG.
FT	DOMAIN	1729	1732	POLY-PRO,	POLY-PRO.
FT	DOMAIN	1741	1744	POLY-ALA,	POLY-ALA.
FT	DOMAIN	1902	1905	POLY-GLY,	POLY-GLY.
FT	DOMAIN	2260	2263	POLY-GLY,	POLY-GLY.
FT	DOMAIN	2404	2407	POLY-SLN,	POLY-SLN.
FT	DOMAIN	2411	2418	POLY-PRO,	POLY-PRO.
FT	DOMAIN	2522	2525	POLY-SER,	POLY-SER.
FT	TYPE	1665	1666	CLEAVAGE BY	CLEAVAGE BY
FT	DISULFID	24	37	BY SIMILARITY.	BY SIMILARITY.
FT	DISULFID	31	46	BY SIMILARITY.	BY SIMILARITY.
FT	DISULFID	48	57	BY SIMILARITY.	BY SIMILARITY.
FT	DISULFID	63	74	BY SIMILARITY.	BY SIMILARITY.
FT	DISULFID	68	87	BY SIMILARITY.	BY SIMILARITY.
FT	DISULFID	89	98	BY SIMILARITY.	BY SIMILARITY.
FT	DISULFID	106	117	BY SIMILARITY.	BY SIMILARITY.
FT	DISULFID	111	127	BY SIMILARITY.	BY SIMILARITY.
FT	DISULFID	129	138	BY SIMILARITY.	BY SIMILARITY.
FT	DISULFID	144	155	BY SIMILARITY.	BY SIMILARITY.
FT	DISULFID	149	164	BY SIMILARITY.	BY SIMILARITY.
FT	DISULFID	166	175	BY SIMILARITY.	BY SIMILARITY.
FT	DISULFID	182	195	BY SIMILARITY.	BY SIMILARITY.
FT	DISULFID	189	204	BY SIMILARITY.	BY SIMILARITY.
FT	DISULFID	206	215	BY SIMILARITY.	BY SIMILARITY.
FT	DISULFID	222	233	BY SIMILARITY.	BY SIMILARITY.
FT	DISULFID	227	243	BY SIMILARITY.	BY SIMILARITY.
FT	DISULFID	245	254	BY SIMILARITY.	BY SIMILARITY.
FT	DISULFID	261	272	BY SIMILARITY.	BY SIMILARITY.
FT	DISULFID	266	281	BY SIMILARITY.	BY SIMILARITY.
FT	DISULFID	283	292	BY SIMILARITY.	BY SIMILARITY.
FT	DISULFID	299	312	BY SIMILARITY.	BY SIMILARITY.
FT	DISULFID	306	321	BY SIMILARITY.	BY SIMILARITY.
FT	DISULFID	323	332	BY SIMILARITY.	BY SIMILARITY.
FT	DISULFID	339	350	BY SIMILARITY.	BY SIMILARITY.

Query Match 15.38; Score 1034.5; DB 1; Length 2556;
 Best Local Similarity 25.88; Pred. No. 3.6e-53;
 Matches 316; Conservative 83; Mismatches 304; Indels 523; Gaps 73;

FT DISULFID 344 359 BY SIMILARITY.
 FT DISULFID 361 370 BY SIMILARITY.
 FT DISULFID 376 387 BY SIMILARITY.
 FT DISULFID 381 398 BY SIMILARITY.

QY 94 CCFPEYEGEMCVPRHCKVCHGR-----IAPNTQCEPFGMGTCSSACDGDH 143
 DB 89 CALGE--SGPLCLPLDACLTPCRNGCTDLTLTEYKCRCPGWSGKSCQQA----- 141
 QY 144 WGPHTSRQCKNGALCPRTIGA--CHCAAGFRGRCE--DRCEOG--TYGNDCHQ- 193
 DB 142 --DPCASN--PCANGGCLPFASATYCHCPSPFHGTGQDYNEGQAKRLCRHGTCNE 198
 QY 194 ---RC-----OQANGC--DHYTGRCRPPYTGAFCE-- 222
 DB 199 VGSRYCVRATHGTGPNCEPRYPVPCSPQNGTCRPTGDTYHACALPGFTGQNCENI 258
 QY 223 DLCPRG--KHGPQ-----EORCP-----CQNGVCCHVTG- 251
 DB 259 DDCGNNCKNGACVGVNTYNCPCPEMTGCTEDVDECQLMPNACQNGTCHNTHG 318
 QY 252 --ECSPGSMGTVCQ-----PCPRGRGKNC--SQEC-- 281
 DB 319 YNCVCVNMTEGDCSENIDDCASACFHGATCHDRVASFYECPRGTGLLCHLNDACIS 378
 QY 282 --QCHNGTCD--ATGQ--CHCSPGYTERCO--DECPVYGVLAETCQCVNGKCY 333
 DB 379 NPCNGSNCDTPNPNKALICTCPSGYTPACSDVDECSLAN-----PCHAGACI 430
 QY 334 HVSGA--CLCAGRAGECEARLCEGLYIKDKRC--PCHLENTHSHPMGE--CA 386
 DB 431 NTLSFEQCLQGYTGPRCEIDV-----NECVSNPC--QNDATCLDQIGEFQCM 477
 QY 387 CKPGMSGLYC-----NE--TCSPGFYGEACQ--QICS--C 415
 DB 478 CMPEGEVHCEVNTDECASSPCLNNGRCLDKINEFQCCPFGTGHLCQDYVDECASTPC 537
 QY 416 QNGADC-----DSV--TGKCTCAPGPKG 436
 DB 538 KNGAKCLDGPMTYTCVCTEGYTGHEVDIDEDCPDCHYSGCKGVATFCLCPGYTG 597
 QY 437 IDCSF-----PCPL-----GTYINS-----SRGCANDAVCS 465
 DB 598 HHCEININESSQPCRLKGTQDPDMAYLCFLAGTTGPNCIINDDCASSPCDSGTCLD 657
 QY 466 PVDD--SCTCKAGMHGVDCSIR-----CPSGTWGFQCNL-----TC----- 499
 DB 658 KIDYECAECERGYTSGKMSNIDECAGNCPHNGSTCEJINGFTCRCEYHDPCLSEV 717
 QY 500 ---OCLNGACNLTLDG--TCTCAPGMRGKCEL----- 527
 DB 718 NECSNPNFVHAGACRDLNGYKCDPDMGSGTGNINNECSNPNVNGTCKDMTSGIYC 777
 QY 528 PCQDGTVLNCAERCD-----CSHADOC--HPTTGH--CRCLPMSGVCHDSV--CARG- 575
 DB 778 TCRBEGFSGPNCQTNINECASNPCLNKGTCLIDVAGYKCNCLLPYTGACVEVLAAPCASP 837
 QY 576 --RMGPNC-----SLPCYC--KNGASCSPDQICECAPGFRGTQRT-----CS 616
 DB 838 CRNGEGEQSEDYESFSCVCTAGAKGCTCEVDINECVLSPCRHAGASCQNTNGXTRCHQ 897
 QY 617 PGFYGHRSQTCPOCVHSSGPCHH-----ITGLDCLPGLFTGALCNE----- 658
 DB 898 AGYSGRNCETDIDDC--RPNCHNGSCTDGINFAFCCLLPGRFRTFCEEDIINECASDPC 955
 QY 659 -----VCSGFRGKRCAG--ICT--CTNNGTCMPIDR--SCQCTPG 693
 DB 956 RANGANTCDVSDYTCCTCPAGFSGIHCENNTDCTESSCGNGTCT--VDGINSFTCLCPG 1013

QY 694 WIGSDC-----SOP-----CPAHMGPNC---IHTCN---CHNGA 722
 DB 1014 FTGSTCQVHVNECBSRFLILGCTCODGRLHRCITPCGYTPKCNQNLVHWCDSPPCKANG 1073
 QY 723 FC-----SAYDECKCTPGMTGLYCTQ----- 744
 DB 1074 KCMQTHRYQ--RCRCPSGMTGLYCDVPSVCEVAARQGVAVARLQHGGLCVADAGNTH 1131
 QY 745 --RCPLEFGYKCALI--CQ-----CQNGADC--DHSIG--QCTRTGFMGRHCDK----- 788
 DB 1132 CRCQAGYTSYCEDLVDECSPPSCQNGATCTDYLGYSCKCVAGHYVNCSEEDCLSH 1191
 QY 789 -----CPSGTGYGCRQICD-----CLNNTSC--DHTG 815
 DB 1192 PCQNGTCLDLPNTYKCSCPGTGVCNEINVDNCPVYDVSHPKCFNNGTCVDQYGG 1251
 QY 816 --TCYSPGWMKARGCDQAGVITVNLN 840
 DB 1252 YSCTCPQFVEGRCE-----GDVN 1270

RESULT 3
 NC1_MOUSE
 ID NC1_MOUSE STANDARD: PRT: 2531 AA.
 AC 001705: Q9QW58: Q99JC2: Q06007: Q61905: Q9R0X7;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Neurogenic locus notch homolog protein 1 precursor (Notch 1) (Motch A) (m114) (P300).
 GN NOTCH1 OR MOTCH.
 OS Mus musculus (mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_Taxid:10090;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC TISSUE=Embryo;
 RX MEDLINE-93194170; PubMed-8449489;
 RA Franco del Amo F., Gendron-Maguire M., Swiatek P.J., Jenkins N.A., Copeland N.G., Gridley T.;
 RT "Cloning, analysis, and chromosomal localization of Notch-1, a mouse homolog of Drosophila Notch.";
 RL Genomics 15:259-264(1993).
 RN [2]
 RP SEQUENCE OF 731-1899 FROM N.A. (ISOFORM 2), AND DEVELOPMENTAL STAGE.
 RC STRAIN=CD-1; TISSUE=Embryo;
 RX MEDLINE-93050801; PubMed-1426644;
 RA Reume A.G., Conlon R.A., Zirnagl R., Yamaguchi T.P., Rossant J.;
 RT "Expression analysis of a Notch homologue in the mouse embryo.";
 RN Dev. Biol. 154:377-387(1992).
 RN [3]
 RP SEQUENCE OF 1551-1647 FROM N.A. (ISOFORM 1), AND DEVELOPMENTAL STAGE.
 RC TISSUE=Embryo;
 RX MEDLINE-93048835; PubMed-1425352;
 RA Franco del Amo F., Smith D.E., Swiatek P.J., Gendron-Maguire M., Greenspan R.J., McMahon A.P., Gridley T.;
 RT "Expression pattern of Motch, a mouse homolog of Drosophila Notch, suggests an important role in early postimplantation mouse development.";
 RN Development 115:737-744(1992).
 RN [4]
 RP SEQUENCE OF 1161-1547 FROM N.A.
 RC STRAIN=C57BL/6 X CBA; TISSUE=Embryo;
 RX MEDLINE-93178563; PubMed-8440332;
 RA Lardelli M., Lendahl U.;
 RT "Motch A and Motch B-two mouse Notch homologues coexpressed in a wide variety of tissues.";
 RN Exp. Cell Res. 204:364-372(1993).
 RN [5]
 RP SEQUENCE OF 1659-1673 FROM N.A.
 RX MEDLINE-99364499; PubMed-10437788;
 RA Lee J.S., Ishimoto A., Yanagawa S.I.;

RT "Murine leukemia provirus-mediated activation of the Notch1 gene leads
RT to induction of HES-1 in a mouse T lymphoma cell line, DL-3."
RL FEBS Lett. 455:276-280(1999).
RN
RN
RN SEQUENCE OF 1950-2201 FROM N.A.
RP MEDLINE-98029496; PubMed-9344671;
RX Messierle M., Folio M., Nehls M., Eggert H., Boehm T.?
RX "Dynamic changes in gene expression during in vitro differentiation of
RT mouse embryonic stem cells";
RL Cytokines Cell. Mol. Ther. 1:139-143(1995).
RN
RN [17]
RP SEQUENCE OF 1655-1659, CLEAVAGE BY FURIN-LIKE CONVERTASE, AND
RX NOTAGENESIS OF 1651-ARG-GLN-ARG-ARG-1654.
RX MEDLINE-98318619; PubMed-9653118;
RA Logeat F., Bessia C., Brou C., Lebail O., Varrault S., Sedah N.G.,
RA Israel A.?
RT "The Notch1 receptor is cleaved constitutively by a furin-like
RT convertase";
RL Proc. Natl. Acad. Sci. U.S.A. 95:8108-8112(1998).
RN [8]
RN PARTIAL SEQUENCE, AND POST-TRANSLATIONAL PROCESSING.
RP MEDLINE-215323956; PubMed-11518718;
RX Saxena M.T., Schroeter E.H., Mumm J.S., Kopan R.?
RA "Murine notch homologs (NL-4) undergo presenilin-dependent
RT proteolysis";
RL J. Biol. Chem. 276:40268-40273(2001).
RN [9]
RN POST-TRANSLATIONAL PROCESSING.
RP MEDLINE-2137376; PubMed-11459941;
RX Mizutani T., Taniguchi Y., Aoki T., Hashimoto N., Hongo T.?
RA "Conservation of the biochemical mechanisms of signal transduction
RT among mammalian Notch family members";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9026-9031(2001).
RN
RN -1- FUNCTION: Functions as a receptor for membrane-bound ligands
CC Jagged2 and Delta to regulate cell-fate determination.
CC Upon ligand activation through the released notch intracellular
CC domain (NICD) it forms a transcriptional activator complex with
CC RBP-1 kappa and activates genes of the enhancer of split locus.
CC Affects the implementation of differentiation, proliferation and
CC apoptotic programs (by similarity). May play an essential role in
CC postimplantation development, probably in some aspect of cell
CC specification and/or differentiation. May be involved in mesoderm
CC development, somit formation and neurogenesis. Involved in the
CC maturation of both CD4+ and CD8+ cells in the thymus.
CC
CC -1- SUBUNIT: Heterodimer of a C-terminal fragment N(NTM) and a N-
CC terminal fragment N(CT) which are probably linked by disulfide
CC bonds.
CC
CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Following
CC proteolytic processing NICD is translocated to the nucleus.
CC
CC -1- ALTERNATIVE PRODUCTS: 2 isoforms: 1 (shown here) and 2; may be
CC produced by alternative splicing.
CC
CC -1- TISSUE SPECIFICITY: Highly expressed in the brain, lung and
CC thymus. Expressed at lower levels in the spleen, bone-marrow,
CC spinal cord, eyes, mammary gland, liver, intestine, skeletal
CC muscle, kidney and heart.
CC
CC -1- DEVELOPMENTAL STAGE: First detected in the mesoderm at 7.5 d.p.c.
CC By 8.5 d.p.c. highly expressed in presomitic mesoderm, mesenchyme
CC and endothelial cells, while much lower levels are seen in the
CC neuroepithelium. Between 9.5-10.5 d.p.c. expressed at high levels
CC in the neuroepithelium. At 13.5 d.p.c. expressed in the surface
CC ectoderm, eye and developing whisker follicles.
CC
CC -1- PTM: Synthesized in the endoplasmic reticulum as an inactive form
CC which is proteolytically cleaved by a furin-like convertase in the
CC trans-golgi network before it reaches the plasma membrane to yield
CC an active ligand-accessible form. Cleavage results in a C-
CC terminal fragment (NTM) and a N-terminal fragment (NCT). Following
CC ligand binding, it is cleaved by TNF-alpha converting enzyme
CC (TACE) to yield a membrane-associated intermediate fragment called
CC notch extracellular truncation (NEXT). This fragment is then
CC cleaved by presenilin dependent gamma-secretase to release a
CC notch-derived peptide containing the intracellular domain (NICD)
CC from the membrane.
CC
CC -1- PTM: Phosphorylated.

CC	-1	SIMILARITY: BELONGS TO THE NOTCH FAMILY.	
CC	-1	SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.	
CC	-1	SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.	
CC	-1	SIMILARITY: CONTAINS 5 ANK REPEATS.	
CC			
CC		This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/or_send_an_email_to_license@isb-sib.ch).	
CC			
DR	EMBL: Z11886;	CAAT794.1.1; -	
DR	EMBL: L02613;	AAK14898.1; -	
DR	EMBL: X68278;	CAA8339.1; -	
DR	EMBL: AJ338029;	CABA0733.1; -	
DR	EMBL: X82562;	CAAT57909.1; -	
DR	HSSP: P00740;	IEDM.	
DR	MGD: MGI:97363;	Notch1.	
DR	InterPro: IPR002110;	ANK.	
DR	InterPro: IPR000152;	Asx_hydroxy1.	
DR	InterPro: IPR000561;	EGF-like.	
DR	InterPro: IPR000742;	EGF_2.	
DR	InterPro: IPR001881;	EGF_Ca.	
DR	InterPro: IPR001438;	EGF_II.	
DR	InterPro: IPR000800;	Notch.	
DR	Pfam: PF00008;	EGF; 35.	
DR	Pfam: PF00023;	ank; 7.	
DR	Pfam: PF00066;	notch; 3.	
DR	PRINTS: PRO0010;	EGFBLOOD.	
DR	PRINTS: PRO1452;	NOTCH.	
DR	SMART: SM00248;	ANK; 3.	
DR	SMART: SM00179;	EGF_CA; 23.	
DR	SMART: SM00001;	EGF_1like; 11.	
DR	SMART: SM00004;	NL; 2.	
DR	PROSITE: PS50088;	ANK_REPEAT; 2.	
DR	PROSITE: PS50297;	ANK_REPEAT_REGION; 1.	
DR	PROSITE: PS00020;	ASX_HYDROXYL; 22.	
DR	PROSITE: PS00022;	EGF_1; 34.	
DR	PROSITE: PS01186;	EGF_2; 27.	
DR	PROSITE: PS01187;	EGF_CA; 21.	
KW	Receptor; Transcription regulation; Activator; Differentiation; Developmental protein; Repeat; ANK repeat; EGF-like domain;		
KW	Transmembrane; Glycoprotein; Signal; Phosphorylation;		
KW	Alternative splicing.		
FT	CHAIN	1	POTENTIAL.
FT	CHAIN	19	NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1.
FT	CHAIN	1711	NORCH EXTRACELLULAR TRUNCATION.
FT	CHAIN	1744	NORCH INTRACELLULAR DOMAIN.
FT	CHAIN	1725	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	1726	POTENTIAL.
FT	DOMAIN	1747	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	20	EGF-LIKE 1.
FT	DOMAIN	59	EGF-LIKE 2.
FT	DOMAIN	102	EGF-LIKE 3.
FT	DOMAIN	140	EGF-LIKE 4.
FT	DOMAIN	178	EGF-LIKE 5.
FT	DOMAIN	218	EGF-LIKE 6.
FT	DOMAIN	257	EGF-LIKE 7.
FT	DOMAIN	295	EGF-LIKE 8.
FT	DOMAIN	333	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	371	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	372	EGF-LIKE 10.
FT	DOMAIN	410	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	412	EGF-LIKE 11.
FT	DOMAIN	450	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	452	EGF-LIKE 12.
FT	DOMAIN	488	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	490	EGF-LIKE 13.
FT	DOMAIN	526	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	528	EGF-LIKE 14.
FT	DOMAIN	564	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	566	EGF-LIKE 15.
FT	DOMAIN	603	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	639	EGF-LIKE 16.
FT	DOMAIN	641	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	676	EGF-LIKE 17.
FT	DOMAIN	714	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	716	EGF-LIKE 18.
FT	DOMAIN	751	CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN	753	EGF-LIKE 19.
FT	DOMAIN		CALCIUM-BINDING (POTENTIAL).

FT DOMAIN 791 827 EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 829 867 EGF-LIKE 22.
 FT DOMAIN 869 905 EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 907 943 EGF-LIKE 24.
 Query Match 15.2%; Score 1028; DB 1; Length 2531;
 Best Local Similarity 25.7%; Pred. No. 8.7e-53;
 Matches 314; Conservative 83; Mismatches 286; Indels 538; Gaps 73;
 86 TMYRRSQCCPFYESEGCYV---PHCADKCVH-GRCI---APNTCCCEPFGMGATNCSS- 137
 121 TLTEKRCRSPGM--SGKSCQADPCASNPANGGCLPESSYICRCPGRHGPTRQD 178
 138 -----ACDGDHWGPHC-----TSRCCCKNALCNP--- 162
 179 VNECSQNPGLCRHGHCHNEISYRCACATHTGPHCELPPYVCPSPSPQONATCRPTD 238
 163 ITGACHCAAGFRGWRCEDECEGTYGNDCHORCOCONGATC-DHY-TGECRCPGYTGA 220
 239 TTHECACLPFGAGNCEAVD-----DCPGR-NCKNGACVGYNTTYKCRCPPEYTGQY 291
 221 C-EDLCPPKHPQCEQRCP--CQNGVCHVYTG--ECSCPSGMMGTVCQ----- 266
 292 CREDV-----DEC-QLMPNACQNAAGTCHNTHGVCVNGWMTGEDCSENIDOCASAA 343
 267 -----PCPEGFRKNC--SQEC---QCNGSTCDA--ATGQ--CHCSFG 301
 344 CQFATGCHDRVAFYCECHGRTGLLCHLKHACISNPCEGSCNCTPNVNGRITCPCPSG 403
 302 YNGERQ---DECPVGYVLCAETQCQVNGKCYHVSQA--CLCBAGPAGERCEARLCP 356
 404 YRGPAQSDVDECDLGAN-----RCHAKKCLNTLGSFECQCLQGTGTCCELDV--- 453
 357 EBLYGKCDKR---PCHLENTSHCPMSGE--CAKPGWSGLYCNET----- 399
 454 -----NECISNPC--QNDATCIDQIGEFQCIOMPVEGYCEINTDECASSPCLHN 502
 400 -----CSPGFGEACQ--QICS--CONGADG--DSVTGCTCAPKRGKD 438
 503 GHCMDKIHFOCCQCRKRGFNGHLQYDVBCEASTPCKNGKKCLDGPNTYTCVTEGTYGH 562
 439 CST---PCPLGT-----YGINCSSRGC---CKNDVACSPVD 468
 563 CEVDIDECDDPCHYSGKCDGVAFTCLQPGYTHHCTNTNNECHSQCRHGRGTQDDD 622
 469 GS--CTCKAGHGVDCSIR-----CPSGTWFGCNITLCOCLNGAGCNTIDG--TCICA 517
 623 NSYLCICLKGTGPNCEIMIDCASNPDCSGT-----CL--DKIDYEACAE 667
 518 PGMREKCEL-----PCQDSTYGLNCAERDCSHADGCHPTT----- 554
 668 PGTGSMCVNINDECAGSPCHNGTCEBDIAGFTC--RC---PEGYHPTCLSEVNECN 721
 555 -----GHCRC-----CLPGWSGVHCD-----SYCAE 574
 722 SNPCIHGACRDLNGYKKCCAPAGWSGTNCDINNCEOSPCVNGGCKMNTSGYCTGAE 781
 575 GWMGPNC-----SLPY-----CKNG 590
 782 GSGSPCOTNINECASNPCLNOGTICIDVAGYKCNCPLEPYTGATGEVVLAPCATSPCKNS 841
 591 ASCSPDDGI---CECAPFGRGTQO-----RIGSGEFG 621
 842 GYCKESEDEYESTCYCPTGMOGTCEVDINCEYKSPCRHAGASCQNTNGSYRCLCAQAGTYG 901
 622 HRCQOTPCQVHSSGPCNH-----ITGLCDCLPGFTGALCNE----- 658
 902 RACESDIDDC--RPNPCHNGSGCTDGTINTAFCDCLPGFGARCEBDINECANPQONGAN 959
 659 -----VPSGRFGKNCAG--ICT--CTNNGTCNPIDR--SCQYPCWIGSD 698
 960 CTDCVDSYTCCTCPVGFNGFHCENMTPTDCESSCFNGGTC--VDGINSFCLCPGPGTGSY 1017

QY 699 C-----SQP-----CPRAHMGPNC---IHTNC---CHNGAF--- 724
 DB 1018 CQYDVEDDSRPLCHGTCQDSYTYKCTCPQGYGLNCONIVRWCDSAPCKNGRCMQT 1077
 QY 725 -SAYDECKCTPWTGLYC---TORCPLEFY--GRDCALICQ----- 760
 DB 1078 NTQY--HCECRSGWTVGNCVDVLSYSCFEVAQAQKRGIDVTLCOHGGLCYDESGKHCHCQA 1135
 QY 761 -----CQNGADC-DHISG--QCTCRFGFMGRHCEQK----- 788
 DB 1136 GYTGSGYCEDEVEDCSFNPCCQONATCTDYLGFSKCVAGYHGSNCSSEINECLSQPCQNG 1195
 QY 789 -----CPSGYTYGCRQICD-----CLNNSNC-DHITG--TEYC 819
 DB 1196 GTCIDLTNSKSCSPRGTVGVCCEINVDCHPPLDPASRSPKCFNNGTCVQDVGGYGTCTC 1255
 QY 820 SPWKGARCDQAGVITVGNLN 840
 DB 1256 PPGVGERCE-----GDVN 1269
 RESULT 4
 NTCL_RAT STANDARD; PRT; 2531 AA.
 AC 007008:
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 15-JUN-1999 (Rel. 38, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Neurogenic locus notch homolog protein 1 precursor (Notch 1).
 GN NOTCH1.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Schwann cell;
 RX MEDLINE=92111383; PubMed=1764995;
 RA Weimaster G., Roberts V.J., Lemke G.;
 RT "A homolog of Drosophila Notch expressed during mammalian
 RT development.";
 RL Development 113:199-205(1991).
 RN [2]
 RP REVISIONS TO 1652-1653.
 RA Weimaster G.;
 RL Submitted (Apr-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP FUNCTION.
 RX MEDLINE=21094508; PubMed=11182080;
 RA Tanigaki K., Nogaki F., Takahashi J., Tashiro K., Kurooka H.,
 RA Honjo T.;
 RT "Notch1 and Notch3 instructively restrict bFGF-responsive multipotent
 RT neural progenitor cells to an astroglial fate.";
 RL Neuron 29:45-55(2001).
 RN [4]
 RP TISSUE SPECIFICITY.
 RX MEDLINE=93202015; PubMed=1295745;
 RA Weimaster G., Roberts V.J., Lemke G.;
 RT "Notch2: a second mammalian Notch gene.";
 RL Development 116:931-941(1992).
 RN [5]
 RP TISSUE SPECIFICITY.
 RX MEDLINE=21331789; PubMed=11438922;
 RA Irvan D.K., Zurcher S.D., Nguyen T., Weimaster G., Kornblum H.I.;
 RT "Expression patterns of Notch1, Notch2, and Notch3 suggest multiple
 RT functional roles for the Notch-Delta signaling system during brain
 RT development.";
 RL J. Comp. Neurol. 436:167-181(2001).
 CC -!- FUNCTION: Functions as a receptor for membrane-bound ligands
 CC Jagged1, Jagged2 and Delta1 to regulate cell-fate determination.
 CC Upon ligand activation through the released notch intracellular
 CC domain (NICD) it forms a transcriptional activator complex with
 CC RBP-J kappa and activates genes of the enhancer of split locus.

FT DISULFID 222 233 BY SIMILARITY.
 FT DISULFID 227 243 BY SIMILARITY.
 FT DISULFID 245 254 BY SIMILARITY.
 FT DISULFID 261 272 BY SIMILARITY.

Query Match 15.2%; Score 1024; DB 1; Length 2531;
 Best Local Similarity 25.8%; Pred. No. 1.5e-52;
 Matches 315; Conservative 83; Mismatches 286; Indels 536; Gaps 74;

QY 86 TMTRRSQQCPGYESEGEMCV---PHCADKCVH-GRCIT---APNTQCEPBGWGTGSS- 137
 DB 121 TLTLEYKRCRPPGW--SGKSCQADPCASNPANGGCLPFESSYICGCPFGHGPCRD 178
 QY 138 -----ACDDHMGPHC-----TSRCKKNGALCNP----- 162
 DB 179 VNECSNPGLCRHGTCHNEIGSYRCACRATHGPHCELPYPFCSPPCQNGTCTPTGD 238
 QY 163 ITGACHCAAGFRGWRCEDEQGTGNDCHORCQNGATC-DHV-TGECRCPPTGTAF 220
 DB 239 TTHECACLPGFAGQNCENVD-----DCPGN-NCRNGACVDGVNTYNCRCPPETGTQY 291
 QY 221 C-EDLCPRKHGRCRQRCP--CQNGVCHHTG--ECCSPSGMGTVCQ----- 266
 DB 292 CTEDV-----DEC-OLMPNACQNGTCHNSHGYNCAVNGWTEDESDNIDCASAA 343
 QY 267 -----PCEGFRFGKNC--SOEC--QCHNGTCTDA-ATGQ--CHCSPG 301
 DB 344 CTGQATCHNRVASFYCECHGRGLLCHLDACISNPEGSCTDPNPNKATICTCPRG 403
 QY 302 YRGERC---DECPVGTGVLCAETGCQVNGKCYHVSAG-CLCBAGFAGERCEARLCP 356
 DB 404 YTGAPASQDVDECALGAN-----PCBHAGCKLNTLISFEQCLQGTGPRCEIDV-- 453
 QY 357 EGIYGIKCKKRC--PCHLEHTHSHCPMSGE--CAKRPMSGLYC----- 396
 DB 454 -----NECISNPC--QNDATCLDQIGEFQICMPAYEGYECINTECASPSCLHN 502
 QY 397 -----NE--TCSPFGEACQ--QICS--CONGADC--DSYTGKCTCAPRGKID 438
 DB 503 GRGVOKINFLGQCPRGFGSHLQDYVDECASTPCKNKAKCLDPRITYCYCEGTGTH 562
 QY 439 CST-----PCPLGTYSINCSSRCCKNDAYCSPVDSCTCKAGW----- 477
 DB 563 CEVDIDECDDPDCHTGL-----CK-DGVAT--FTCLQDPYTHHCETINECH 608
 QY 478 -----HGVDCSTR-----CPSGTWFGCINLTCQ-----CLNGACMTLDG-TCTCAP 518
 DB 609 SOPCRHGTGQDNDNTYLCCLKLGTGPNCETINLDDCASNPDCSGTCLDKIDYECACEP 668
 QY 519 GMRGKCEL-----PCDDGTGVLNCAERCDSCSHADGCHPTT----- 554
 DB 669 GTTGSNCNNNIDECAGSPHNGTCEJDAGFTC-RC--PEGHDTCLSEVNECHS 722
 QY 555 -----GHCR-----CLPGNSGVHCD-----SYCAEG 575
 DB 723 NPICHGACRDGLNGYKCDAPGMSGTNCIDINNNECESNPVNGTCKDMTSGYVTCRBS 782
 QY 576 RMGPNC-----SLRCP-----CKNKA 591
 DB 783 FSGPNCQTNINECASNPCLNQTCLIDVAGYKCNCLPYTGATCEVYLAFCATSPCKNSG 842
 QY 592 SCSPDDGI-----CECAPGFRGTTCQ-----RISCPFYH 622
 DB 843 VKKESDYSFSCVCTPGMGQGTCEIDINECYKSPCRHASCQNTGSTRCLQAGTYGR 902
 QY 623 RSCQTCPOCVHSSGPCNH-----ITGLCDCLPFTGALCNE----- 658
 DB 903 NCESDIDDC--RNPCHNGSGCTDGVNAAFCDCLPFGCAFCEEDINECATNPQNGANG 960
 QY 659 -----VCPBGRFGKACAG--ICT--CTNNGCTNPIDR--SCQCTPGWTGSDC 699
 DB 961 TDCVDSYTCCTPTGFGNIGHCENNTPDCTESSCFNGTGC--VDGINSFTCLCPGFGTGYC 1018

QY 700 -----SOP-----CPPAHMGPN--IHTCN--CHNGAFC----- 724
 DB 1019 QYDVNCEBDRPCLHSGTCDSDYGTCTKCTCPGCTGLNCLNLYRMCDASACKNGCKMQTN 1078
 QY 725 SAYDECKCTPGWTGLYC--TORCPLGFY--GRDICALCQ----- 760
 DB 1079 TQY--HCECRSGWTGNCVDVLSVSCVAAQKRIGIDVTLTQHGGLCVDEEDKHYCHCQAG 1136
 QY 761 -----CONGADC-DHISG-OCTCTGTGMRHCEBK----- 788
 DB 1137 YTGSGYCEDEVDRCSPNPCCNGATCTDYLGFSCKVAGYHGSNCSEINECLSQPCQNG 1196
 QY 789 -----CPSGTYGCGRQLCD-----CLNNGTC-DHITG--TCYCS 820
 DB 1197 TCIDLTNTYKSCSPKRTQSVHCEINVDCHPPLDPASRSPKCFNNCTCYDQGGTYCTCP 1256
 QY 821 PGMKARGCDQAGYIIVGNLN 840
 DB 1257 PGFVGRCR-----GDVN 1269

RESULT 5

NTC2_HUMAN
 ID NTC2_HUMAN STANDARD; PRT; 2471 AA.
 AC 004721; Q9H240; Q99734;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Neutrogenic locus notch homolog protein 2 precursor (Notch 2) (hn2).
 GN NOTCH2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RC TISSUE=Brain;
 RA Blaumüller C.M., Mann R.S.;
 RT "Complete human notch 2 (hn2) cDNA sequence."
 RL Submitted (NOV-2000) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RC TISSUE=FROM N.A.
 RA Cornea R.G., Camargo A.A., Moreira E.S., Simpson A.J.;
 RT "Human Notch2, a novel member of cell-fate determining NOTCH
 family."
 RL Submitted (OCT-2000) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RC TISSUE=T-cell;
 RA Lemasson I., Devaux C., Mesnard J.M.;
 RT "Partial sequence of EGF-like repeat domain of human Notch2 mRNA."
 RL Submitted (NOV-1996) to the EMBL/Genbank/DBJ databases.
 RN [4]
 RC TISSUE=Brain;
 RA MEDLINE=93265135; PubMed=9244302;
 RA Stifani S., Blaumüller C.M., Redhead N.J., Hill R.E.,
 RT "Human homologs of a Drosophila enhancer of split gene product define
 a novel family of nuclear proteins."
 RL Nat. Genet. 2:119-127(1992).
 RN [5]
 RC POST-TRANSLATIONAL PROCESSING.
 RA MEDLINE=97386453; PubMed=9244302;
 RA Blaumüller C.M., Qi H., Zagouras P., Artavanis-Tsakonas S.;
 RT "Intracellular cleavage of Notch leads to a heterodimeric receptor on
 the plasma membrane."
 RL Cell 90:281-291(1997).
 RN [6]
 RC IDENTIFICATION OF LIGANDS.
 RA MEDLINE=99180765; PubMed=10079256;
 RX Gray G.E., Mann R.S., Mitsiadis E., Henrique D., Carcangiu M.-L.,

RA Banks A., Lelman J., Ward D., Ish-Horowitz D., Artavanis-Tsakonas S.;
 RT "Human ligands of the Notch receptor";
 RL Am. J. Pathol. 154:785-794(1999).
 CC -1- FUNCTION: Functions as a receptor for membrane-bound ligands.
 CC Jagged1, Jagged2 and Deltal to regulate cell-fate determination.
 CC Upon ligand activation through the released notch intracellular
 CC domain (NICD) it forms a transcriptional activator complex with
 CC RBP-J kappa and activates genes of the enhancer of split locus.
 CC Affects the implementation of differentiation, proliferation and
 CC apoptotic programs (By similarity).
 CC -1- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
 CC terminal fragment N(EC) which are probably linked by disulfide
 CC bonds (By similarity).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Following
 CC proteolytical processing NICD is translocated to the nucleus.
 CC -1- TISSUE SPECIFICITY: Expressed in the brain, heart, kidney, lung,
 CC skeletal muscle and liver.
 CC -1- PTM: Synthesized in the endoplasmic reticulum as an inactive form
 CC which is proteolytically cleaved by a furin-like convertase in the
 CC trans-Golgi network before it reaches the plasma membrane to yield
 CC an active, ligand-accessible form. Cleavage results in a C-
 CC terminal fragment N(TM) and a N-terminal fragment N(EC). Following
 CC ligand binding, it is cleaved by TNF-alpha converting enzyme
 CC (TACE) to yield a membrane-associated intermediate fragment called
 CC notch extracellular truncation (NEXT). This fragment is then
 CC cleaved by presenilin dependent gamma-secretase to release a
 CC notch-derived peptide containing the intracellular domain (NICD)
 CC from the membrane (By similarity).
 CC -1- PTM: Phosphorylated (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE NOTCH FAMILY.
 CC -1- SIMILARITY: CONTAINS 35 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 LIN/NOTCH REPEATS.
 CC -1- SIMILARITY: CONTAINS 6 ANK REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib.ch).
 CC -----
 CC DR EMBL: AF308601; AAA36377.2; -;
 CC DR EMBL: AF315356; ANG37073.1; -;
 CC DR EMBL: U77493; AAB19224.1; -;
 CC DR HSSP: P00740; IEDM.
 CC DR GeneW: HGNC:7882; NOTCH2.
 CC DR MIM: 600275; -;
 CC DR InterPro: IPR002110; ANK.
 CC DR InterPro: IPR000152; ASX_hydroxyl.
 CC DR InterPro: IPR000361; EGF-like.
 CC DR InterPro: IPR000742; EGF-2.
 CC DR InterPro: IPR001881; EGF_Ca.
 CC DR InterPro: IPR001438; EGF_II.
 CC DR InterPro: IPR002049; Lamlnln_EGF.
 CC DR InterPro: IPR000800; Notch.
 CC DR Pfam: PF00008; EGF; 35.
 CC DR Pfam: PF00023; ank; 6.
 CC DR Pfam: PF00066; notch; 2.
 CC DR PRINTS: PRO0010; EGFBL00.
 CC DR PRINTS: PRO0011; EGFAMININ.
 CC DR SMART: SM00246; ANK; 4.
 CC DR SMART: SM00179; EGF_Ca; 22.
 CC DR SMART: SM00001; EGF_Like; 12.
 CC DR SMART: SM00004; NL; 2.
 CC DR PROSITE: PS50088; ANK_REPEAT; 4.
 CC DR PROSITE: PS50297; ANK_REPEAT_REGION; 1.
 CC DR PROSITE: PS00010; ASX_HYDROXYL; 22.
 CC DR PROSITE: PS00022; EGF-1; 34.
 CC DR PROSITE: PS01186; EGF-2; 29.
 CC DR PROSITE: PS01187; EGF_Ca; 22.
 CC DR Receptor: Transcription regulation: Activator: Differentiation;
 CC Developmental protein: Repeat; ANK repeat; EGF-like domain;
 CC KW

KW Transmembrane; Glycoprotein; Signal; Phosphorylation.
 FT SIGNAL 1 25
 FT CHAIN 26 2471
 FT CHAIN 1666 2471
 FT CHAIN 1697 2471
 FT CHAIN 26 1677
 FT TRANSMEM 1678 1698
 FT DOMAIN 1699 2471
 FT DOMAIN 26 63
 FT DOMAIN 64 102
 FT DOMAIN 105 143
 FT DOMAIN 144 180
 FT DOMAIN 182 219
 FT DOMAIN 221 258
 FT DOMAIN 260 296
 FT DOMAIN 298 336
 FT DOMAIN 338 374
 FT DOMAIN 375 413
 FT DOMAIN 415 454
 FT DOMAIN 456 494
 FT DOMAIN 494 530
 FT DOMAIN 532 568
 FT DOMAIN 570 605
 FT DOMAIN 607 643
 FT DOMAIN 645 680
 FT DOMAIN 682 718
 FT DOMAIN 720 755
 FT DOMAIN 757 793
 FT DOMAIN 795 831
 FT DOMAIN 833 871
 FT DOMAIN 873 909
 FT DOMAIN 911 947
 FT DOMAIN 948 985
 FT DOMAIN 987 1023
 FT DOMAIN 1025 1061
 FT DOMAIN 1063 1099
 FT DOMAIN 1101 1147
 FT DOMAIN 1149 1185
 FT DOMAIN 1187 1223
 FT DOMAIN 1225 1262
 FT DOMAIN 1264 1302
 FT DOMAIN 1304 1343
 FT DOMAIN 1374 1412
 FT REPEAT 1420 1456
 FT REPEAT 1456 1503
 FT REPEAT 1503 1535
 FT REPEAT 1535 1571
 FT REPEAT 1571 1607
 FT REPEAT 1607 1643
 FT REPEAT 1643 1679
 FT REPEAT 1679 1715
 FT REPEAT 1715 1751
 FT REPEAT 1751 1787
 FT REPEAT 1787 1823
 FT REPEAT 1823 1859
 FT REPEAT 1859 1895
 FT REPEAT 1895 1931
 FT REPEAT 1931 1967
 FT REPEAT 1967 2003
 FT REPEAT 2003 2038
 FT DOMAIN 1645 1681
 FT DOMAIN 1681 1717
 FT DOMAIN 1717 1753
 FT DOMAIN 1753 1789
 FT DOMAIN 1789 1825
 FT DOMAIN 1825 1861
 FT DOMAIN 1861 1897
 FT DOMAIN 1897 1933
 FT DOMAIN 1933 1969
 FT DOMAIN 1969 2005
 FT DOMAIN 2005 2041
 FT DOMAIN 2041 2077
 FT DOMAIN 2077 2113
 FT DOMAIN 2113 2149
 FT DOMAIN 2149 2185
 FT DOMAIN 2185 2221
 FT DOMAIN 2221 2257
 FT DOMAIN 2257 2293
 FT DOMAIN 2293 2329
 FT DOMAIN 2329 2365
 FT DOMAIN 2365 2401
 FT DOMAIN 2401 2437
 FT DOMAIN 2437 2473
 FT DOMAIN 2473 2509
 FT DOMAIN 2509 2545
 FT DOMAIN 2545 2581
 FT DOMAIN 2581 2617
 FT DOMAIN 2617 2653
 FT DOMAIN 2653 2689
 FT DOMAIN 2689 2725
 FT DOMAIN 2725 2761
 FT DOMAIN 2761 2797
 FT DOMAIN 2797 2833
 FT DOMAIN 2833 2869
 FT DOMAIN 2869 2905
 FT DOMAIN 2905 2941
 FT DOMAIN 2941 2977
 FT DOMAIN 2977 3013
 FT DOMAIN 3013 3049
 FT DOMAIN 3049 3085
 FT DOMAIN 3085 3121
 FT DOMAIN 3121 3157
 FT DOMAIN 3157 3193
 FT DOMAIN 3193 3229
 FT DOMAIN 3229 3265
 FT DOMAIN 3265 3301
 FT DOMAIN 3301 3337
 FT DOMAIN 3337 3373
 FT DOMAIN 3373 3409
 FT DOMAIN 3409 3445
 FT DOMAIN 3445 3481
 FT DOMAIN 3481 3517
 FT DOMAIN 3517 3553
 FT DOMAIN 3553 3589
 FT DOMAIN 3589 3625
 FT DOMAIN 3625 3661
 FT DOMAIN 3661 3697
 FT DOMAIN 3697 3733
 FT DOMAIN 3733 3769
 FT DOMAIN 3769 3805
 FT DOMAIN 3805 3841
 FT DOMAIN 3841 3877
 FT DOMAIN 3877 3913
 FT DOMAIN 3913 3949
 FT DOMAIN 3949 3985
 FT DOMAIN 3985 4021
 FT DOMAIN 4021 4057
 FT DOMAIN 4057 4093
 FT DOMAIN 4093 4129
 FT DOMAIN 4129 4165
 FT DOMAIN 4165 4201
 FT DOMAIN 4201 4237
 FT DOMAIN 4237 4273
 FT DOMAIN 4273 4309
 FT DOMAIN 4309 4345
 FT DOMAIN 4345 4381
 FT DOMAIN 4381 4417
 FT DOMAIN 4417 4453
 FT DOMAIN 4453 4489
 FT DOMAIN 4489 4525
 FT DOMAIN 4525 4561
 FT DOMAIN 4561 4597
 FT DOMAIN 4597 4633
 FT DOMAIN 4633 4669
 FT DOMAIN 4669 4705
 FT DOMAIN 4705 4741
 FT DOMAIN 4741 4777
 FT DOMAIN 4777 4813
 FT DOMAIN 4813 4849
 FT DOMAIN 4849 4885
 FT DOMAIN 4885 4921
 FT DOMAIN 4921 4957
 FT DOMAIN 4957 4993
 FT DOMAIN 4993 5029
 FT DOMAIN 5029 5065
 FT DOMAIN 5065 5101
 FT DOMAIN 5101 5137
 FT DOMAIN 5137 5173
 FT DOMAIN 5173 5209
 FT DOMAIN 5209 5245
 FT DOMAIN 5245 5281
 FT DOMAIN 5281 5317
 FT DOMAIN 5317 5353
 FT DOMAIN 5353 5389
 FT DOMAIN 5389 5425
 FT DOMAIN 5425 5461
 FT DOMAIN 5461 5497
 FT DOMAIN 5497 5533
 FT DOMAIN 5533 5569
 FT DOMAIN 5569 5605
 FT DOMAIN 5605 5641
 FT DOMAIN 5641 5677
 FT DOMAIN 5677 5713
 FT DOMAIN 5713 5749
 FT DOMAIN 5749 5785
 FT DOMAIN 5785 5821
 FT DOMAIN 5821 5857
 FT DOMAIN 5857 5893
 FT DOMAIN 5893 5929
 FT DOMAIN 5929 5965
 FT DOMAIN 5965 6001
 FT DOMAIN 6001 6037
 FT DOMAIN 6037 6073
 FT DOMAIN 6073 6109
 FT DOMAIN 6109 6145
 FT DOMAIN 6145 6181
 FT DOMAIN 6181 6217
 FT DOMAIN 6217 6253
 FT DOMAIN 6253 6289
 FT DOMAIN 6289 6325
 FT DOMAIN 6325 6361
 FT DOMAIN 6361 6397
 FT DOMAIN 6397 6433
 FT DOMAIN 6433 6469
 FT DOMAIN 6469 6505
 FT DOMAIN 6505 6541
 FT DOMAIN 6541 6577
 FT DOMAIN 6577 6613
 FT DOMAIN 6613 6649
 FT DOMAIN 6649 6685
 FT DOMAIN 6685 6721
 FT DOMAIN 6721 6757
 FT DOMAIN 6757 6793
 FT DOMAIN 6793 6829
 FT DOMAIN 6829 6865
 FT DOMAIN 6865 6901
 FT DOMAIN 6901 6937
 FT DOMAIN 6937 6973
 FT DOMAIN 6973 7009
 FT DOMAIN 7009 7045
 FT DOMAIN 7045 7081
 FT DOMAIN 7081 7117
 FT DOMAIN 7117 7153
 FT DOMAIN 7153 7189
 FT DOMAIN 7189 7225
 FT DOMAIN 7225 7261
 FT DOMAIN 7261 7297
 FT DOMAIN 7297 7333
 FT DOMAIN 7333 7369
 FT DOMAIN 7369 7405
 FT DOMAIN 7405 7441
 FT DOMAIN 7441 7477
 FT DOMAIN 7477 7513
 FT DOMAIN 7513 7549
 FT DOMAIN 7549 7585
 FT DOMAIN 7585 7621
 FT DOMAIN 7621 7657
 FT DOMAIN 7657 7693
 FT DOMAIN 7693 7729
 FT DOMAIN 7729 7765
 FT DOMAIN 7765 7801
 FT DOMAIN 7801 7837
 FT DOMAIN 7837 7873
 FT DOMAIN 7873 7909
 FT DOMAIN 7909 7945
 FT DOMAIN 7945 7981
 FT DOMAIN 7981 8017
 FT DOMAIN 8017 8053
 FT DOMAIN 8053 8089
 FT DOMAIN 8089 8125
 FT DOMAIN 8125 8161
 FT DOMAIN 8161 8197
 FT DOMAIN 8197 8233
 FT DOMAIN 8233 8269
 FT DOMAIN 8269 8305
 FT DOMAIN 8305 8341
 FT DOMAIN 8341 8377
 FT DOMAIN 8377 8413
 FT DOMAIN 8413 8449
 FT DOMAIN 8449 8485
 FT DOMAIN 8485 8521
 FT DOMAIN 8521 8557
 FT DOMAIN 8557 8593
 FT DOMAIN 8593 8629
 FT DOMAIN 8629 8665
 FT DOMAIN 8665 8701
 FT DOMAIN 8701 8737
 FT DOMAIN 8737 8773
 FT DOMAIN 8773 8809
 FT DOMAIN 8809 8845
 FT DOMAIN 8845 8881
 FT DOMAIN 8881 8917
 FT DOMAIN 8917 8953
 FT DOMAIN 8953 8989
 FT DOMAIN 8989 9025
 FT DOMAIN 9025 9061
 FT DOMAIN 9061 9097
 FT DOMAIN 9097 9133
 FT DOMAIN 9133 9169
 FT DOMAIN 9169 9205
 FT DOMAIN 9205 9241
 FT DOMAIN 9241 9277
 FT DOMAIN 9277 9313
 FT DOMAIN 9313 9349
 FT DOMAIN 9349 9385
 FT DOMAIN 9385 9421
 FT DOMAIN 9421 9457
 FT DOMAIN 9457 9493
 FT DOMAIN 9493 9529
 FT DOMAIN 9529 9565
 FT DOMAIN 9565 9601
 FT DOMAIN 9601 9637
 FT DOMAIN 9637 9673
 FT DOMAIN 9673 9709
 FT DOMAIN 9709 9745
 FT DOMAIN 9745 9781
 FT DOMAIN 9781 9817
 FT DOMAIN 9817 9853
 FT DOMAIN 9853 9889
 FT DOMAIN 9889 9925
 FT DOMAIN 9925 9961
 FT DOMAIN 9961 9997
 FT DOMAIN 9997 10033
 FT DOMAIN 10033 10069
 FT DOMAIN 10069 10105
 FT DOMAIN 10105 10141
 FT DOMAIN 10141 10177
 FT DOMAIN 10177 10213
 FT DOMAIN 10213 10249
 FT DOMAIN 10249 10285
 FT DOMAIN 10285 10321
 FT DOMAIN 10321 10357
 FT DOMAIN 10357 10393
 FT DOMAIN 10393 10429
 FT DOMAIN 10429 10465
 FT DOMAIN 10465 10501
 FT DOMAIN 10501 10537
 FT DOMAIN 10537 10573
 FT DOMAIN 10573 10609
 FT DOMAIN 10609 10645
 FT DOMAIN 10645 10681
 FT DOMAIN 10681 10717
 FT DOMAIN 10717 10753
 FT DOMAIN 10753 10789
 FT DOMAIN 10789 10825
 FT DOMAIN 10825 10861
 FT DOMAIN 10861 10897
 FT DOMAIN 10897 10933
 FT DOMAIN 10933 10969
 FT DOMAIN 10969 11005
 FT DOMAIN 11005 11041
 FT DOMAIN 11041 11077
 FT DOMAIN 11077 11113
 FT DOMAIN 11113 11149
 FT DOMAIN 11149 11185
 FT DOMAIN 11185 11221
 FT DOMAIN 11221 11257
 FT DOMAIN 11257 11293
 FT DOMAIN 11293 11329
 FT DOMAIN 11329 11365
 FT DOMAIN 11365 11401
 FT DOMAIN 11401 11437
 FT DOMAIN 11437 11473
 FT DOMAIN 11473 11509
 FT DOMAIN 11509 11545
 FT DOMAIN 11545 11581
 FT DOMAIN 11581 11617
 FT DOMAIN 11617 11653
 FT DOMAIN 11653 11689
 FT DOMAIN 11689 11725
 FT DOMAIN 11725 11761
 FT DOMAIN 11761 11797
 FT DOMAIN 11797 11833
 FT DOMAIN 11833 11869
 FT DOMAIN 11869 11905
 FT DOMAIN 11905 11941
 FT DOMAIN 11941 11977
 FT DOMAIN 11977 12013
 FT DOMAIN 12013 12049
 FT DOMAIN 12049 12085
 FT DOMAIN 12085 12121
 FT DOMAIN 12121 12157
 FT DOMAIN 12157 12193
 FT DOMAIN 12193 12229
 FT DOMAIN 12229 12265
 FT DOMAIN 12265 12301
 FT DOMAIN 12301 12337
 FT DOMAIN 12337 12373
 FT DOMAIN 12373 12409
 FT DOMAIN 12409 12445
 FT DOMAIN 12445 12481
 FT DOMAIN 12481 12517
 FT DOMAIN 12517 12553
 FT DOMAIN 12553 12589
 FT DOMAIN 12589 12625
 FT DOMAIN 12625 12661
 FT DOMAIN 12661 12697
 FT DOMAIN 12697 12733
 FT DOMAIN 12733 12769
 FT DOMAIN 12769 12805
 FT DOMAIN 12805 12841
 FT DOMAIN 12841 12877
 FT DOMAIN 12877 12913
 FT DOMAIN 12913 12949
 FT DOMAIN 12949 12985
 FT DOMAIN 12985 13021
 FT DOMAIN 13021 13057
 FT DOMAIN 13057 13093
 FT DOMAIN 13093 13129
 FT DOMAIN 13129 13165
 FT DOMAIN 13165 13201
 FT DOMAIN 13201 13237
 FT DOMAIN 13237 13273
 FT DOMAIN 13273 13309
 FT DOMAIN 13309 13345
 FT DOMAIN 13345 13381
 FT DOMAIN 13381 13417
 FT DOMAIN 13417 13453
 FT DOMAIN 13453 13489
 FT DOMAIN 13489 13525
 FT DOMAIN 13525 13561
 FT DOMAIN 13561 13597
 FT DOMAIN 13597 13633
 FT DOMAIN 13633 13669
 FT DOMAIN 13669 13705
 FT DOMAIN 13705 13741
 FT DOMAIN 13741 13777
 FT DOMAIN 13777 13813
 FT DOMAIN 13813 13849
 FT DOMAIN 13849 13885
 FT DOMAIN 13885 13921
 FT DOMAIN 13921 13957
 FT DOMAIN 13957 14000
 FT DOMAIN 14000 14043
 FT DOMAIN 14043 14086
 FT DOMAIN 14086 14129
 FT DOMAIN 14129 14172
 FT DOMAIN 14172 14215
 FT DOMAIN 14215 14258
 FT DOMAIN 14258 14301
 FT DOMAIN 14301 14344
 FT DOMAIN 14344 14387
 FT DOMAIN 14387 14430
 FT DOMAIN 14430 14473
 FT DOMAIN 14473 14516
 FT DOMAIN 14516 14559
 FT DOMAIN 14559 14602
 FT DOMAIN 14602 14645
 FT DOMAIN 14645 14688
 FT DOMAIN 14688 14731
 FT DOMAIN 14731 14774
 FT DOMAIN 14774 14817
 FT DOMAIN 14817 14860
 FT DOMAIN 14860 14903
 FT DOMAIN 14903 14946
 FT DOMAIN 14946 14989
 FT DOMAIN 14989 15032
 FT DOMAIN 15032 15075
 FT DOMAIN 15075 15118
 FT DOMAIN 15118 15161
 FT DOMAIN 15161 15204
 FT DOMAIN 15204 15247
 FT DOMAIN 15247 15290
 FT DOMAIN 15290 15333
 FT DOMAIN 15333 15376
 FT DOMAIN 15376 15419
 FT DOMAIN 15419 15462
 FT DOMAIN 15462 15505
 FT DOMAIN 15505 15548
 FT DOMAIN 15548 15591
 FT DOMAIN 15591 15634
 FT DOMAIN 15634 15677
 FT DOMAIN 15677 15720
 FT DOMAIN 15720 15763
 FT DOMAIN 15763 15806
 FT DOMAIN 15806 15849
 FT DOMAIN 15849 15892
 FT DOMAIN 15892 15935
 FT DOMAIN 15935 15978
 FT DOMAIN 15978 16021
 FT DOMAIN 16021 16064
 FT DOMAIN 16064 16107
 FT DOMAIN 16107 16150
 FT DOMAIN 16150 16193
 FT DOMAIN 16193 16236
 FT DOMAIN 16236 16279
 FT DOMAIN 16279 16322
 FT DOMAIN 16322 16365
 FT DOMAIN 16365 16408
 FT DOMAIN 16408 16451
 FT DOMAIN 16451 16494
 FT DOMAIN 16494 16537
 FT DOMAIN 16537 16580
 FT DOMAIN 16580 16623
 FT DOMAIN 16623 16666
 FT DOMAIN 16666 16709
 FT DOMAIN 16709 16752
 FT DOMAIN 16752 16795
 FT DOMAIN 16795 16838
 FT DOMAIN 16838 16881
 FT DOMAIN 16881 16924
 FT DOMAIN 16924 16967
 FT DOMAIN 16967 17010
 FT DOMAIN 17010 17053
 FT DOMAIN 17053 17096
 FT DOMAIN 17096 17139
 FT DOMAIN 17139 17182
 FT DOMAIN 17182 17225
 FT DOMAIN 17225 17268
 FT DOMAIN 17268 17311
 FT DOMAIN 17311 17354
 FT DOMAIN 17354 17397
 FT DOMAIN 17397 17440
 FT DOMAIN 17440 17483
 FT DOMAIN 17483 17526
 FT DOMAIN 17526 17569
 FT DOMAIN 17569 17612
 FT DOMAIN 17612 17655
 FT DOMAIN 17655 17698
 FT DOMAIN 17698 17741
 FT DOMAIN 17741 17784
 FT DOMAIN 17784 17827
 FT DOMAIN 17827 17870
 FT DOMAIN 17870 17913
 FT DOMAIN 17913 17956
 FT DOMAIN 17956 17999
 FT DOMAIN 17999 18042
 FT DOMAIN 18042 18085
 FT DOMAIN 18085 18128
 FT DOMAIN 18128 18171
 FT DOMAIN 18171 18214
 FT DOMAIN 18214 18257
 FT DOMAIN 18257 18300
 FT DOMAIN 18300 18343
 FT DOMAIN 18343 18386
 FT DOMAIN 18386 18429
 FT DOMAIN 18429 18472
 FT DOMAIN 18472 18515
 FT DOMAIN 18515 18558
 FT DOMAIN 18558 18601
 FT DOMAIN 18601 18644
 FT DOMAIN 18644 18687
 FT DOMAIN 18687 18730
 FT DOMAIN 18730 18773
 FT DOMAIN 18773 18816
 FT DOMAIN 18816 18859
 FT DOMAIN 18859 18902
 FT DOMAIN 18902 18945
 FT DOMAIN 18945 18988
 FT DOMAIN 18988 19031
 FT DOMAIN 19031 19074
 FT DOMAIN 19074 19117
 FT DOMAIN 19117 19160
 FT DOMAIN 19160 19203
 FT DOMAIN 19203 19246
 FT DOMAIN 19246 19289
 FT DOMAIN 19289 19332
 FT DOMAIN 19332 19375
 FT DOMAIN 19375 19418
 FT DOMAIN 19418 19461
 FT DOMAIN 19461 19504
 FT DOMAIN 19504 19547
 FT DOMAIN 19547 19590
 FT DOMAIN 19590 19633
 FT DOMAIN 19633 19676
 FT DOMAIN 19676 19719
 FT DOMAIN 19719 19762
 FT DOMAIN 19762 19805
 FT DOMAIN 19805 19848
 FT DOMAIN 19848 19891
 FT DOMAIN 19891 19934
 FT DOMAIN 19934 19977
 FT DOMAIN 19977 20020
 FT DOMAIN 20020 20063
 FT DOMAIN 20063 20106
 FT DOMAIN 20106 20149
 FT DOMAIN 20149 20192
 FT DOMAIN 20192 20235
 FT DOMAIN 20235 20278
 FT DOMAIN 20278 20321
 FT DOMAIN 20321 20364
 FT DOMAIN 20364 20407
 FT DOMAIN 20407 20450
 FT DOMAIN 20450 20493
 FT DOMAIN 20493 20536
 FT DOMAIN 20536 20579
 FT DOMAIN 20579 20622
 FT DOMAIN 20622 20665
 FT DOMAIN 20665 20708
 FT DOMAIN 20708 20751
 FT DOMAIN 20751 20794
 FT DOMAIN 20794 20837
 FT DOMAIN 20837 20880
 FT DOMAIN 20880 20923
 FT DOMAIN 20923 20966
 FT DOMAIN 20966 21009
 FT DOMAIN 21009 21052
 FT DOMAIN 21052 21095
 FT DOMAIN 21095 21138
 FT DOMAIN 21138 21181
 FT DOMAIN 21181 21224
 FT DOMAIN 21224 21267
 FT DOMAIN 21267 21310
 FT DOMAIN 21310 21353
 FT DOMAIN 21353 21396
 FT DOMAIN 21396 21439
 FT DOMAIN 21439 21482
 FT DOMAIN 21482 21525
 FT DOMAIN 21525 21568
 FT DOMAIN 21568 21611
 FT DOMAIN 21611 21654
 FT DOMAIN 21654 21697
 FT DOMAIN 21697 21740
 FT DOMAIN 21740 21783
 FT DOMAIN 21783 21826
 FT DOMAIN 21826 21869
 FT DOMAIN 21869 21912
 FT DOMAIN 21912 21955
 FT DOMAIN 21955 21998
 FT DOMAIN 21998 22041
 FT DOMAIN 22041 22084
 FT DOMAIN 22084 22127
 FT DOMAIN 22127 22170
 FT DOMAIN 22170 22213
 FT DOMAIN 22213 22256
 FT DOMAIN 22256 22299
 FT DOMAIN 22299 22342
 FT DOMAIN 22342 22385
 FT DOMAIN 22385 22428
 FT DOMAIN 22428 22471
 FT DOMAIN 22471 22514
 FT DOMAIN 22514 22557
 FT DOMAIN 22557 22600
 FT DOMAIN 22600 22643
 FT DOMAIN 22643 22686
 FT DOMAIN 22686 22729
 FT DOMAIN 22729 22772
 FT DOMAIN 22772 22815
 FT DOMAIN 22815 22858
 FT DOMAIN 22858 22901
 FT DOMAIN 22901 22944
 FT DOMAIN 22944 22987
 FT DOMAIN 22987 23030
 FT DOMAIN 23030 23073
 FT DOMAIN 23073 23116
 FT DOMAIN 23116 23159
 FT DOMAIN 23159 23202
 FT DOMAIN 23202 23245
 FT DOMAIN 23245 23288
 FT DOMAIN 23288 23331
 FT DOMAIN 23331 23374
 FT DOMAIN 23374 23417
 FT DOMAIN 23417 23460
 FT DOMAIN 23460 23503
 FT DOMAIN 23503 23546
 FT DOMAIN 23546 23589
 FT DOMAIN 23589 23632
 FT

FT DISULFID 248 257 BY SIMILARITY.
 FT DISULFID 264 275 BY SIMILARITY.
 FT DISULFID 269 284 BY SIMILARITY.
 FT DISULFID 286 295 BY SIMILARITY.

Query Match 15.0%; Score 1014.5; DB 1; Length 2471;
 Best Local Similarity 23.0%; Pred. No. 5.2e-52;
 Matches 348; Conservative 106; Mismatches 354; Indels 707; Gaps 78;

QY 93 OCCPGFYESGEMCVPHCADKCYHGR-CLAPNTCQ-----CEPGNGTNCSSACDG 141
 DB 91 RCASGR--TGEDCOYSTHPCFVSRLCLNGTCHMLSRDYECTCQVFTGKREC----- 142
 QY 142 DHMGPHCTSRCCCKNGKALCNPTIG--ACHCAGFRGMRCEDECEGTYNDCDHQRCQCN 199
 DB 143 -QMTDACLH-PCANGSTCTTYANQFSCKCLTGFYGKCE-----TDVNECDIPGHCH 194
 QY 200 GATCDHYTG-ECRCPPGYTGAFCEDL-----CPGKHG 231
 DB 195 GGTCLNLPSTYQCCPQGTGQCDLYVCAAPSPCVNGTCRQNGDFTFECNCLPFGEG 254
 QY 232 PCECO-----RCP-----CONGV 245
 DB 255 STCERNIDCPNHRCONGVGVVNTYNCRCRPOWTGQCTEDYDECLQPNACQNGT 314
 QY 246 CHHYTG--ECSPSGMGTVCQ-----HCPGRGKNC- 277
 DB 315 CANRNGYGCYVNGSRGSDDCSENIDDCAFASCTPSTCIDYVASSCMCPKAGLCH 374
 QY 278 -SOEC--QCHNGTCD-ATGQ-CHCSPGYTERCO--DECPVGYVLCATCQC 326
 DB 375 LDDACISNCHKALCDTNPNGOYICTPQGYKADCTEDYDECAM-----ANSNC 427
 QY 327 VNGGKCYHSGA--CLCEGFAGERGE-----ARLCEGLX 360
 DB 428 EHAGCQVNTDGAHCECLKYGAPCEMDINECHSDPCQNDATCLDKIGTCLCMPGR 487
 QY 361 GIKCDR-----CP-----CHLE-----NTHSC 378
 DB 488 GHCELEINECOSNFCVNNQCVDAVNRQCLCPREFTGPVQIIDDCSSPCLNGANC 547
 QY 379 --HPSGECACKPGMSGLYCNET-----CSPGEYGEAC-QQ 411
 DB 548 IDHPNGYEGOCATGFTGYLCEENINDCPDPCHHGQCCDGIDSYICINPRTGMAICSDQ 607
 QY 412 I-----CSCQCA-----DCDS--VTG-----K 427
 DB 608 IDECYSSPCLNDGRCIDLVNGYQCNQCPETSGVNCENINFDCAASNPCILGICMDGINRYS 667
 QY 428 CTCAGFKGIDC-----STPCPLGTGINCSS--RCGCK-----NDAVCS 465
 DB 668 CTCSEFTGQRNIDIDECASNPCRKATCINGVNGFRICPEGRPHHPSCYQVNECLSN 727
 QY 466 P-VDSCCT-----CKAGHGVDCST-----RCPST 490
 DB 728 PCIHGNCITGLSGYKCLDAGWVINGCEVDKNECLSNPCQNGTCDNLVNGYRCCKKGF 787
 QY 491 WFGGCLTQ-----CLNGGAC-----507
 DB 788 KQYNCQVNIDECASNPCLNQGTGCFDIDISYTHCVLPYTKNCQVLYAPCSNPENAAV 847
 QY 508 -----NTLDGTCAPAGNGEKCCLPCODGTGLNCAERCDSDSHDGGPTTGH--CRL 560
 DB 848 CKEPNFESTYCLCAPAGMOGOKCTIDIDE-----CISK-PCMHGCLINTGOSTWCECP 900
 QY 561 PMSGVHCDVCAEGRWGPNCSLPYCKNGKASCSPDDGI-----CECAPGFRGTTCOR-- 613
 DB 901 PEFSGMDCEBIDIDCLANP-----CQNGSGCM--DGVNTPSCCLPLPFTDKQCTDMN 951
 QY 614 -----ICSPFYGHRCSCOTPCQVSS-----GPCHHITGL-- 644
 DB 952 ECLSEPCKNGTCDYVNSYTKCOAGFPGVHCENNINCECTESSCFNGTCT--VVGINSF 1009

QY 645 -CDCLPGFTGALC-----NEV-----CPSGFRKNC--AGICT- 674
 DB 1010 SCLCPVGFPGSCFLHEINCCSSHPCLINEGTCDVGLATGYRSCPLGTGKNCQTVNLCSR 1069
 QY 675 --CTNNGTC--NPIDRSCCYPPWIGSDCSQP----- 702
 DB 1070 SPCKMKGTGVOKKASQCLCPSGMAGAYCDVNVSCDIAASRGVLEHLQHSQVCINA 1129
 QY 703 -----CPPAMWPNCC--HTC--NCHNAPFSAYDC--ECKTPEWTLXTQR-- 745
 DB 1130 GNTHYCOCPLYGTGYSCEBOLDECAASNPCOHGATCDSFTGGRCECVPGYQVNGEYVD 1189
 QY 746 -----CPLGEYG-----KCALICQONQADC-DHISQ- 773
 DB 1190 ECQNPQNGTICIDLVNHFKSCPPGTGGLCEENIDDCAPRCLNGGCKMDRIGYS 1249
 QY 774 -----CTCRTGFMGRHCEQKCPST 793
 DB 1250 CRLPGFAGERCEGDINECLSNPSCSESLDCTQLTNDYLCVCRSAFTGRHCE-----T 1303
 QY 794 YGTCRQICDCLANNSTCHITG-----TCYSPKMGKARGDAGVLYVNLNLSRTSTA 848
 DB 1304 FVDVCPQM-PCLNGGTCAVASMPPGFLCRCPGFSGARGQSS--CGVRC----- 1351
 QY 849 LPADSYQIAGIILVLYVFLFLATFIYHKKQKKE-----SSMPATYTPARVYNA 904
 DB 1352 -----KRGEQVYTFASPR-CFQSPRSDCS 1376
 QY 905 DYTIGTLPHSNGNANSHYFTNPSTHTLQCATSPHVNRDMVTYTKSKNQLFVNLEKN 964
 DB 1377 GCASS--PCOHGSGC--HPQRPPYYS-CQCA-PFSGSR-----CELYT 1415
 QY 965 VNGRGRPYGDCGTGLPDMKHGTYLNEAGAGLDRSYGKSLKDLGKNS--EYNSNC 1021
 DB 1416 APPST--PPATLSOYCAPKARDGVCE-----ACNSHACQMDGGDC 1455
 QY 1022 SLSSENPATYTIKDP 1036
 DB 1456 SL-TMENPWANCSSP 1469

RESULT 6
 NTC2_RAT
 ID NTC2_RAT STANDARD: PRT: 2471 AA.
 AC Q9QW30.
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Neurogenic locus notch homolog protein 2 precursor (Notch 2).
 GN NOTCH2.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Rattus.
 OX NCBI_Taxid=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=93202015; PubMed=1295745;
 RA Weinmaster G., Roberts V.J., Lemke G.,
 RT "Notch2: a second mammalian Notch gene.";
 RL Development 116:931-941(1992).
 RN [2]
 RP TISSUE SPECIFICITY.
 RX MEDLINE=21331789; PubMed=11438922;
 RA Irvan D.K., Zurcher S.D., Nguyen T., Weinmaster G., Kornblum H.I.;
 RT "Expression patterns of Notch1, Notch2, and Notch3 suggest multiple
 functional roles for the Notch-Delta signaling system during brain
 development.";
 RL J. Comp. Neurol. 436:167-181(2001).
 CC -I- FUNCTION: Functions as a receptor for membrane-bound ligands
 CC Jagged2 and Delta1 to regulate cell-fate determination.
 CC upon ligand activation through the released notch intracellular
 CC domain (NICD) it forms a transcriptional activator complex with

FT DISULFID 269 284 BY SIMILARITY.
 FT DISULFID 286 295 BY SIMILARITY.
 FT DISULFID 302 315 BY SIMILARITY.
 FT DISULFID 309 324 BY SIMILARITY.
 FT DISULFID 326 335 BY SIMILARITY.
 FT DISULFID 342 353 BY SIMILARITY.
 FT DISULFID 347 362 BY SIMILARITY.
 FT DISULFID 364 373 BY SIMILARITY.
 FT DISULFID 379 390 BY SIMILARITY.
 FT DISULFID 384 401 BY SIMILARITY.
 FT DISULFID 403 412 BY SIMILARITY.
 FT DISULFID 419 433 BY SIMILARITY.
 FT DISULFID 427 442 BY SIMILARITY.
 FT DISULFID 444 453 BY SIMILARITY.
 FT DISULFID 460 471 BY SIMILARITY.
 FT DISULFID 465 480 BY SIMILARITY.
 FT DISULFID 482 491 BY SIMILARITY.
 FT DISULFID 498 509 BY SIMILARITY.
 FT DISULFID 503 518 BY SIMILARITY.
 FT DISULFID 520 529 BY SIMILARITY.
 FT DISULFID 536 547 BY SIMILARITY.
 FT DISULFID 541 556 BY SIMILARITY.
 FT DISULFID 558 567 BY SIMILARITY.
 FT DISULFID 574 584 BY SIMILARITY.

Query Match 14.88; Score 998; DB 1; Length 2471;
 Best Local Similarity 24.48; Pred. No. 4; Be-51;
 Matches 321; Conservative 79; Mismatches 322; Indels 596; Gaps 70;

QY 17 CHWIGTASPLNLEDPNVCNWSYVTVQESY-----PHPEQIYVTSCTDIL 64
 DB 121 CHMS-----MDTECTQVGTGKQOCWTDVCLSHPEN--GSTCSVA 163
 QY 65 NWFCSTRHRYSTRAYRKHGETMYRRKSQCCPFGYSEEMKVPKCAKCVGRCLAPNTC 124
 DB 164 NQFSC-----RCPAGI--TGOKCADINECDIGRQOHGCTC 198
 QY 125 QCEBGMGTNCSACDGDHMGPHCTS-----RCQKNGALC--NPITGACHCAAPRG 175
 DB 199 LNLF--GSYRQ--CPQFTQOHCDSPYPCAPSPCVNNGVGTCTROTGDTSCHCLPREG 254
 QY 176 WRCEDEBQGTGYNDCHQRCQONGANC-DHY-TGECRCPPGYTGAF-C-EDLCPPGKHGP 232
 DB 255 SNCRNID-----DCPRNH-KQONGVCVDGVNTYNCRCPPMTGQFCTEDV-----D 300
 QY 233 QC-BQRCQONGVCHNHTG--ECSQPSGMMGTVCQGP----- 267
 DB 301 ECLLDPNACQNGGCTTNNGGYGCVNWSGDDCSENIDDCAFASCTPGSTCIDRVASF 360
 QY 268 ---CPEGRFGKNC--SQEC--QCHNGGTCA--ATGQ--CHCSPGYGERQ--DECP 312
 DB 361 SCLPREGAGILCHLDADACISNPKHAKALCDTNPLNGYICTCTQATYAGACTEDYDCA 420
 QY 313 VGTGVILCAETQCCVNGKCYHVSQA--CLCEAGFAGERCE----- 351
 DB 421 M-----ANSNPEHAKCVNTDGAFCCEIKGYAGRCEMDINECHSDPCQNDATCLD 473
 QY 352 ---ARLCPREGVIGIKDKR-----CP-----CHLE- 373
 DB 474 KIGGFTCLMGPFGVHCELEVNQCQSNPCVNNQCVDKVRFOCLCPFGTGPVCOQIDI 533
 QY 374 ---NTHSC--HPMSGECAKCPMSGLYCNET----- 399
 DB 534 DDCSTPLNKAACIDHNGEYECOCANGFTTCLDENIDNDPPRCHHGQODGIDSTYC 593
 QY 400 -CSFGFTGEAC-QQITCSC-----QNGADCSTVYG-KCTCAPGFGKIDC-----STP 442
 DB 594 ICNPGYMGALCSQIDDEYSSPCINDGRCIDLVNGYQCNQCGPSTGLNCEINFDCASNP 653
 QY 443 CPLGTY--GIN-----CS-----SRGC-----CKNAV----- 464
 DB 654 CLHAGACVDGINRYSVCVSPGFTGRCNCIDIDEACASNPCKRDKATCINDVNGFRCMCPREGPH 713

QY 465 -----SP-YDGSCT-----CKAGMHGVDCSI----- 484
 DB 714 HPSCYQVNECLSPSCINGNCTGLSGYKCLADAGWGINEVAKNECLSPQONGTCN 773
 QY 485 -----RCPSTGWFGSCNLTQO-----CLNGAC----- 507
 DB 774 NLVNGYRCTCKKGFYKQVNIDECASNPCLNGTCLIDVSGYTCHEMLPYTGKNCQTV 833
 QY 508 -----NLDGTCAPGMREKCELPQODGYGLNCAERDCSHAD 548
 DB 834 LAPSPNCEANAAYCKEAPNESEFTCLAPGMOQRCYVDVDE-----CVSK-PCMMNG 886
 QY 549 GCHPYTGH--CRCLPGMSGVHCDVSCAEGRMGPMCSLPYCKNGASC--SPDDGICECAP 604
 DB 887 ICHNTQSYMECECPFGSGMCEB-----DINDCLANPQONGSCYDKYNTFSCLCLP 939
 QY 605 GFRGTTQOR-----ICSPFYGHRCSTQPCQVHSS----- 635
 DB 940 GFVDDKCTDNECLSEPCKNGTCSDVYNSYTCPCPAGFHGHCENNIDECTESSCFNG 999
 QY 636 GRCHHITGL-----GDCLEPFTGALC-----NE-----VQPSGRFG 666
 DB 1000 GTC--VDSINSFSCLPYGTGFPCLADINECSSNPCLNSTCYDGLCTYCTCPLGTYG 1057
 QY 667 KNC---AGICT---CTNNGTC--NPIDRSCQCYPGMIGSDCSQ----- 701
 DB 1058 KNCQTLVNLCSPPSCKNKNGTQAGKARPRCLCPPGMOAYCYDVNLVSCKAALQKPYVE 1117
 QY 702 -----PCPPAHMGPNC---HTC---NCHNAGFCSAYDG--ECKTP 735
 DB 1118 HLCQHSIGICINAGMWHQCQCLGYTGSYCEBQDLDEACASNPQGHATCSDFIGYRCEVP 1177
 QY 736 GWTLGYCTQR-----CPGFTG-----KCALICOQON 763
 DB 1178 GYGAVNCEYEVEDQONQPCQNGTCTIDLNVHFKSCPPGTGGLCEENIDDCAPAPCLN 1237
 QY 764 GADC-DHTSG----- 781
 DB 1238 GGQCVDRIGGYSKCLRPFAGERCEBDINECLSPCSSEGLDCLQKNNQCYCRASFT 1297
 QY 782 GRHCE--QKCPSTGYGRCQICDCLNNSYCDHTGT-----CYSPGWMKARGCDOA 831
 DB 1298 GRHCEFTLVDYCPQR-----PCLNGTGCAVANSNPDGFCRCPPGFGSARGQSS 1345

RESULT 7
 NTC2_MOUSE
 ID NTC2_MOUSE STANDARD; PRT; 2470 AA.
 AC O35516; 060941; 006008;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Neurogenic locus notch homolog protein 2 precursor (Notch 2) (Notch B).
 GN NOTCH2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=Thymus;
 RA Hamada Y., Higuchi M., Tsujimoto Y.;
 RT "Complete amino acid sequence and multiform transcripts encoded by a single copy of mouse Notch2 gene.";
 RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.
 RP SEQUENCE OF 316-1518 FROM N.A.
 RC STRAIN=C57BL/6 X CBA; TISSUE=embryo;
 RX MEDLINE=93178563; PubMed=8440332;
 RA Lardelli M., Lendahl U.;
 RT "Notch A and Notch B-two mouse Notch homologues coexpressed in a wide variety of tissues.";

RL Exp. Cell Res. 204:364-372(1993).
 [3]
 RP SEQUENCE OF 1765-2153 FROM N.A.
 RX MEDLINE-97075110; PubMed-8917536;
 RA Milner L.A., Bigas A., Kopan R., Brashem-Stein C., Bernstein I.D.,
 RT Martin D.I.;
 "Inhibition of granulocytic differentiation by mNotch1.";
 RL Proc. Natl. Acad. Sci. U.S.A. 93:13014-13019(1996).
 RN [4]
 RP FUNCTION.
 RX MEDLINE-99396706; PubMed-10393120;
 RA Hamada Y., Kadohawa Y., Okabe M., Ikawa M., Coleman J.R.,
 RA Tsujimoto Y.;
 RT "Mutation in ankyrin repeats of the mouse Notch2 gene induces early
 embryonic lethality.";
 RL Development 126:3415-3424(1999).
 RN [5]
 RP DEVELOPMENTAL STAGE, AND ALTERNATIVE SPLICING.
 RX MEDLINE-95333693; PubMed-7609614;
 RA Higuchi M., Kiyama H., Hayakawa T., Hamada Y., Tsujimoto Y.;
 RT "Differential expression of Notch1 and Notch2 in developing and adult
 mouse brain.";
 RL Brain Res. Mol. Brain Res. 29:263-272(1995).
 RN [6]
 RP POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS OF MET-1699.
 RX MEDLINE-21523956; PubMed-11518718;
 RA Saxena M.T., Schroeter E.H., Mumm J.S., Kopan R.;
 RT "Murine notch homologs (NL-4) undergo presenilin-dependent
 proteolysis.";
 RL J. Biol. Chem. 276:40268-40273(2001).
 RN [7]
 RP POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS OF MET-1699.
 RX MEDLINE-21374376; PubMed-11459941;
 RA Mizutani T., Taniguchi Y., Aoki T., Hashimoto N., Honjo T.;
 RT "Conservation of the biochemical mechanisms of signal transduction
 among mammalian Notch family members.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9026-9031(2001).
 CC -1- FUNCTION: Functions as a receptor for membrane-bound ligands
 Jagged2, Jagged2 and Delta to regulate cell-fate determination.
 Upon ligand activation through the released notch intracellular
 domain (NICD) it forms a transcriptional activator complex with
 RBP-J kappa and activates genes of the enhancer of split locus.
 Affects the implementation of differentiation, proliferation and
 apoptotic programs (By similarity). May play an essential role in
 postimplantation development, probably in some aspect of cell
 specification and/or differentiation.
 CC -1- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
 terminal fragment N(EC) which are probably linked by disulfide
 bonds.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Following
 proteolytical processing NICD is translocated to the nucleus.
 CC -1- ALTERNATIVE PRODUCTS: 2 isoforms: 1 (shown here) and 2; may be
 produced by alternative splicing.
 CC -1- TISSUE SPECIFICITY: Expressed in the brain, liver, kidney,
 neuroepithelia, somites, optic vesicles and branchial arches, but
 not heart.
 CC -1- DEVELOPMENTAL STAGE: Expressed in the embryonic ventricular zone,
 the postnatal epidermal cells, and the choroid plexus throughout
 embryonic and postnatal development.
 CC -1- PTM: Synthesized in the endoplasmic reticulum as an inactive form
 which is proteolytically cleaved by a furin-like convertase in the
 trans-Golgi network before it reaches the plasma membrane to yield
 an active, ligand-accessible form. Cleavage results in a C-
 terminal fragment N(TM) and a N-terminal fragment N(EC). Following
 ligand binding, it is cleaved by TNF-alpha converting enzyme
 (TACE) to yield a membrane-associated intermediate fragment called
 notch extracellular truncation (NEXT). This fragment is then
 cleaved by presenilin dependent gamma-secretase to release a
 notch-derived peptide containing the intracellular domain (NICD)
 from the membrane.
 CC -1- PTM: Phosphorylated.
 CC -1- SIMILARITY: BELONGS TO THE NOTCH FAMILY.
 CC -1- SIMILARITY: CONTAINS 34.5 EGF-LIKE DOMAINS.

CC -1- SIMILARITY: CONTAINS 2 LIN/NOTCH REPEATS.
 CC -1- SIMILARITY: CONTAINS 6 ANK REPEATS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by, and for commercial
 entities requires a license agreement (See <http://www.isb-sb.ch/announce/>
 or send an email to license@isb-sb.ch).
 CC
 CC EMBL: D32210; BAA22094.1; -
 CC EMBL: X68279; CAA48340.1; -
 CC EMBL: U31881; AAC52924.1; -
 CC HSSP: P61009; IFSB.
 CC MGD: MGI:97364; Notch2.
 CC InterPro: IPR002110; ANK.
 CC InterPro: IPR000152; Asx_hydroxyl.
 CC InterPro: IPR000561; EGF-like.
 CC InterPro: IPR000742; EGF_2.
 CC InterPro: IPR001881; EGF_Ca.
 CC InterPro: IPR001438; EGF_II.
 CC InterPro: IPR000800; Notch.
 CC Pfam: PF00008; EGF_35.
 CC Pfam: PF00023; ank; 6.
 CC Pfam: PF00066; notch; 2.
 CC PRINTS: PR00010; EGFLOOD.
 CC PRINTS: PR01452; NOTCH.
 CC SMART: SM00248; ANK; 4.
 CC SMART: SM00179; EGF_CA; 22.
 CC SMART: SM00001; EGF_Like; 12.
 CC SMART: SM00004; NL; 3.
 CC PROSITE: PS50088; ANK_REPEAT; 4.
 CC PROSITE: PS50297; ANK_REPEAT_REGION; 1.
 CC PROSITE: PS00010; ASX_HYDROXYL; 22.
 CC PROSITE: PS00022; EGF_1; 33.
 CC PROSITE: PS01186; EGF_2; 27.
 CC PROSITE: PS01187; EGF_CA; 22.
 CC Receptor: Transcription regulation; Activator; Differentiation;
 CC Developmental protein; Repeat; ANK repeat; EGF-like domain;
 CC Transmembrane; Glycoprotein; Signal; Phosphorylation;
 CC Alternative splicing.
 CC SIGNAL 1 25
 CC FT CHAIN 26 2470
 CC FT CHAIN 1666 2470
 CC FT CHAIN 1697 2470
 CC FT CHAIN 26 1677
 CC FT TRANSMEM 1678 1698
 CC FT DOMAIN 1699 2470
 CC FT DOMAIN 26 63
 CC FT DOMAIN 64 102
 CC FT DOMAIN 105 143
 CC FT DOMAIN 144 180
 CC FT DOMAIN 182 219
 CC FT DOMAIN 221 256
 CC FT DOMAIN 258 294
 CC FT DOMAIN 296 334
 CC FT DOMAIN 336 372
 CC FT DOMAIN 373 411
 CC FT DOMAIN 413 452
 CC FT DOMAIN 454 490
 CC FT DOMAIN 492 528
 CC FT DOMAIN 530 566
 CC FT DOMAIN 568 603
 CC FT DOMAIN 603 641
 CC FT DOMAIN 643 678
 CC FT DOMAIN 680 716
 CC FT DOMAIN 718 753
 CC FT DOMAIN 755 791
 CC FT DOMAIN 793 829
 CC FT DOMAIN 831 869
 CC FT DOMAIN 871 907
 CC FT DOMAIN 909 945
 CC
 CC POTENTIAL.
 CC NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 2.
 CC NOTCH EXTRACELLULAR TRUNCATION.
 CC NOTCH INTRACELLULAR DOMAIN.
 CC EXTRACELLULAR (POTENTIAL).
 CC POTENTIAL.
 CC CYTOPLASMIC (POTENTIAL).
 CC EGF-Like 1.
 CC EGF-Like 2.
 CC EGF-Like 3.
 CC EGF-Like 4.
 CC EGF-Like 5.
 CC EGF-Like 6 (INCOMPLETE).
 CC EGF-Like 7.
 CC EGF-Like 8.
 CC EGF-Like 9.
 CC EGF-Like 10.
 CC EGF-Like 11.
 CC EGF-Like 12.
 CC EGF-Like 13.
 CC EGF-Like 14.
 CC EGF-Like 15.
 CC EGF-Like 16.
 CC EGF-Like 17.
 CC EGF-Like 18.
 CC EGF-Like 19.
 CC EGF-Like 20.
 CC EGF-Like 21.
 CC EGF-Like 22.
 CC EGF-Like 23.
 CC EGF-Like 24.

```

FT DOMAIN 947 983 EGF-LIKE 25, CALCIIUM-BINDING (POTENTIAL)
FT DOMAIN 985 1021 EGF-LIKE 26, CALCIIUM-BINDING (POTENTIAL)
FT DOMAIN 1023 1059 EGF-LIKE 27, CALCIIUM-BINDING (POTENTIAL)
FT DOMAIN 1061 1097 EGF-LIKE 28,
FT DOMAIN 1099 1145 EGF-LIKE 29,
FT DOMAIN 1147 1183 EGF-LIKE 30,
FT DOMAIN 1185 1221 EGF-LIKE 31, CALCIIUM-BINDING (POTENTIAL)
FT DOMAIN 1223 1260 EGF-LIKE 32, CALCIIUM-BINDING (POTENTIAL)
FT DOMAIN 1262 1300 EGF-LIKE 33,
FT DOMAIN 1302 1345 EGF-LIKE 34,
FT DOMAIN 1372 1410 EGF-LIKE 35,
FT REPEAT 1418 1454 LIN/NOTCH 2.
FT REPEAT 1501 1533 LIN/NOTCH 1.
FT REPEAT 1825 1869 ANK 1.
FT REPEAT 1874 1903 ANK 2.
FT REPEAT 1907 1937 ANK 3.
FT REPEAT 1941 1970 ANK 4.
FT REPEAT 1974 2003 ANK 5.
FT REPEAT 2007 2036 ANK 6.
FT DOMAIN 1645 1648 POLY-ALA.
FT DOMAIN 1992 1995 POLY-LEU.
FT DOMAIN 2183 2189 POLY-SER.
FT DOMAIN 2423 2428 POLY-GLY.
FT DOMAIN 2445 2450 BY SIMILARITY.
FT DISULFID 28 41 BY SIMILARITY.
FT DISULFID 35 51 BY SIMILARITY.
FT DISULFID 53 62 BY SIMILARITY.
FT DISULFID 79 79 BY SIMILARITY.

```

Query Match 14.7%; Score 993; DB 1; Length 2470;
 Best Local Similarity 25.8%; Pred. No. 9,5e-51;
 Matches 316; Conservative 87; Mismatches 295; Indels 526; Gaps 71;

```

QY 93 QCCGFESGMCYPHADKCVHR-CIAPITCO-----CEGWSGTNCSACDG 141
DB 91 RCAGGF--TGEDCQYSTSHPCFVSRCQNGTCHMLSDYECTCYGFTGKQC----- 142
QY 142 DHMPCHTSRCCNGKALCNPIITG--ACHCAAGFRGWRCE-----DRCEO----- 184
DB 143 -QWMDALSL-PCENGSTCTSVASQFSCKCPAGLTGOKCEADINECDIPGRCKQNGGTCLN 200
QY 185 -GTYGNDG-----HQRCC-----CONCATDHY--TGECRPPGYTGAFC--D 223
DB 201 LPSGYRCQCGGFTGQCDSPYVRLPCVNGGTCTGTGDFLFCNCLPGEFGSTCERNID 260
QY 224 LCPRGK--HQPCCQ-----RCP-----CONGVCHHTG-- 251
DB 261 DCPHNKCONGVGVGVNTYNCRCPPQMTGQCTEDYDECLLPNACQNGGTCTNRNGY 320
QY 252 ECSCPSGMGTVCQGP-----CPGRFRGKNC--SQEC-- 281
DB 321 GCVCVNMSSGDDCESENIDDCAYASCTPGSTCIDRVASFSLCPRGKAGLCHLDADATSN 380
QY 282 QCHNGTCD--ATGQ--CHCSPGYTGERCC--DECPVGTGYVLAETCCQVNGRCYH 334
DB 381 PCHGALCDITPLNGOYITCPCQYKGADEEDVDECAM--ANSNCPCEHAKCVN 433
QY 335 VSQA--CLCEAGFGERCE-----ARLCEGLYGTCKDKR- 367
DB 434 TDGAFHCECLKGYAGPRCEMDINECHSDPCONDATCLDKTIGFTCLCMPGKGVHCELEV 493
QY 368 -----CP-----CHLE-----NTHSC--HPMSGE 384
DB 494 NEQSNPCVNNQCVDVYNNRQCICPRGFTGVPQOIIDDCSSITPLNGAKCIDHPNGYE 553
QY 385 CACPRGNSGLYCNET-----CSPGYGAC--QOI----- 412
DB 554 CQCATGFTGLCDENINCPDPCHGQCQDGLDYTCICNPGYMGALICSPQIDECYSSP 613
QY 413 -----CSCQNA-----DCDS-----VTG-----KTCGARGFK 435
DB 614 CLNDRCITDLYNGYQCNCPGTSGLNCINEFDCAINPCMHGVCVDINRYSCVSPGFT 673

```

```

QY 436 GIDC-----SNPCPLGYTGINCSS--RCGCK-----NDVACSP-VDSGCT 472
DB 674 GQRCNIDIDECASNPCKKATCTINDVNGFRICPEGHPHPSYQVNECLSPCTHNGCT 733
QY 473 -----CRAGHGVDCSI-----RCPSGTWFGGCLNT 498
DB 734 GGLSGYICLDAGVWGVNCEVDKNECLSNPCQNGGTGNLNVNGYRCCKGFKYCNQVN 793
QY 499 CQ-----CLNGAC-----NTL 510
DB 794 IDECASNPCLNOGTCFDDVSGYTCCHMLPYTGKNCQTVLAPCSNPENNAVCKEADNFE 853
QY 511 DGTCTAGPNGRGEKCELPDQDGTGYGLNCAERDCSHADGCHPTGH--CRCLPMSGVHC 568
DB 854 SFSCLCAPNGQKRCYTDVDE-----CLSK-PCMNNGVCHNTGSGYVCECPFGSGMDC 906
QY 569 DSVCAEGRMGPNCSLPYCKRNGASCSPDDGI--CECARGFTTCQR----- 613
DB 907 EE-----DINDLANPCQNGGSCV--DHVNFSCCHGFGIDKQOTDMNECLSPCK 957
QY 614 -----ICSPGYGHRCSQTCPCQVHSS--GPHHTITGL----CDCLPGF 651
DB 958 NGTCSDYVNSYTCTCPAGFHGVHCENNIDECTESSCFNGGTC--VDGINSFSLCPVGF 1015
QY 652 TGAIC-----NE-----VCPSGRFGKNC--AGICT--CTNNGT 680
DB 1016 TGPRCLDINDIECSSNPCLNAGTCVDGLGTYRCITPLGTGKNCQTVLNLCSRSCKRKGT 1075
QY 681 C-NPDRSCQCYPGWIGSDC--SQPCPPA--HMGPNCTHTCNCHNGAFCSAYD-GECK 732
DB 1076 CVGKARPHCLCPGMDGAYCDVLYNSCKAALQKVVEHILQ-HSGICINAGTTHHQ 1134
QY 733 CTBPWTGLYCTQRPLGFYFKDCLLICOQNGADC-DHIS-QCTCTGFMRGRCBK-- 788
DB 1135 CPLGTSYCEEQI-----DECA-SNPCQHGATCNFISGIRCEVPGVGVCEYEVD 1187
QY 789 -----CPSGTYGCGRQICD-----CLNNSTC-DHNG-T 816
DB 1188 EQQNPQONGTCTIDLYNHFFKSCPPRTGLCESENIDECAGSRHCLNGQCVDRIGYT 1247
QY 817 CYCSPGKARGCDQAGYIIVGNLN 840
DB 1248 CRCLPGEFAGERCE-----GDIN 1264

RESULT 8
NTCL_BRARE
ID NTCL_BRARE STANDARD; PRT; 2437 AA.
AC P46530;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Neurogenic locus notch homolog protein 1 precursor.
GN NOTCH1 OR NOTCH.
OS Brachydanio rerio (zebrafish) (Danio rerio)..
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7935;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=94128602; PubMed=8297791;
RA Bierkamp C.; Campos-Ortega J.A.;
RT "A zebrafish homologue of the Drosophila neurogenic gene Notch and
its pattern of transcription during early embryogenesis.";
RL Mech. Dev. 43:87-100(1993).
CC -!- FUNCTION: IMPLICATED IN CELL FATE SPECIFICATIONS DURING
EMBRYO DEVELOPMENT. MAY BE INVOLVED IN THE FORMATION OF THE
CC NEURAL PLATE, NOTOCHORD AND BRAIN VESICLES.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- DEVELOPMENTAL STAGE: Expressed in all cells in pregastrulation
stages. During gastrulation is differentially expressed,

```



```

FT DISULFID 627 636 BY SIMILARITY.
FT DISULFID 643 653 BY SIMILARITY.
FT DISULFID 648 662 BY SIMILARITY.
FT DISULFID 664 673 BY SIMILARITY.
FT DISULFID 680 691 BY SIMILARITY.
FT DISULFID 685 700 BY SIMILARITY.
FT DISULFID 702 711 BY SIMILARITY.
FT DISULFID 718 728 BY SIMILARITY.
FT DISULFID 723 737 BY SIMILARITY.
FT DISULFID 739 748 BY SIMILARITY.
FT DISULFID 755 766 BY SIMILARITY.
FT DISULFID 760 775 BY SIMILARITY.
FT DISULFID 777 786 BY SIMILARITY.
FT DISULFID 793 804 BY SIMILARITY.
FT DISULFID 798 813 BY SIMILARITY.
FT DISULFID 815 824 BY SIMILARITY.
FT DISULFID 831 842 BY SIMILARITY.
FT DISULFID 836 853 BY SIMILARITY.
FT DISULFID 855 864 BY SIMILARITY.
FT DISULFID 871 882 BY SIMILARITY.
FT DISULFID 876 891 BY SIMILARITY.
FT DISULFID 893 902 BY SIMILARITY.
FT DISULFID 909 920 BY SIMILARITY.
FT DISULFID 914 929 BY SIMILARITY.
FT DISULFID 931 940 BY SIMILARITY.
FT DISULFID 947 958 BY SIMILARITY.
FT DISULFID 952 967 BY SIMILARITY.
FT DISULFID 969 978 BY SIMILARITY.

```

Query Match 14.68; Score 987; DB 1; Length 2437;
 Best Local Similarity 24.88; Pred. No. 2,1e-50;
 Matches 310; Conservative 81; Mismatches 320; Indels 538; Gaps 70;

```

QY 91 KSCCPGFESEGMCPVPCADKCVHRC-----IAINTQCEEGNGTNSACD 140
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 85 KDCVLAG--SDRLCLFVPHNACMSPCRNGTSLTLDTFTCCQPGWSGKTCTQLA-- 140

QY 141 GDHMGPHCTSRCCCKNGALCNPTIG--ACHCAAGFRGRCEBRCBGTGNDCH--ORCQC 197
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 141 -----DPCASN--PCANGGQCAFESHYICTCPNPHGOTCTQD--NECANSPSPC 188

QY 198 QNGATCDHVTGE--CRCPGTGAFCEDL-----CPPGK 229
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 189 RNGGCIINEVSGSYLCRCPPEYTGPHCQRLXQPLPSPCRSGGTQVSTDTHTSCILPBF 248

QY 230 HGPOCE-----QRC-----PCONG 243
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 249 TGQTCENHNVDCCTQHACENGSPCIDGINTVNCODKHMTGQCTEDVDECELSPNACONG 308

QY 244 GVCHHVTG--ECSPSPGMMGTVCQ-----PCPEGRFGKN 276
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 309 GTCHNTIGFHCVCVNGMTGSDGCSBNIDDCASACSHGATCHDRVASFCECPHRTGLL 368

QY 277 C--SOEC--OCHNGGTCA--ATGQ--CHCSPGTGERCQ--DECPVGTGYVLCATGC 324
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 369 CHLDACISNCCOKGSMNDTNPVSGKALCTCPRGYTSACNMQDIDECISGAN----- 420

QY 325 QCVGKCYHVSAG--CLCEAGFAGEBCEARLCPBGGLGICDKRCRCPHLNTHSCHMS 382
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 421 PCHGGRCRLNKGSCFQCCLOGYBSPCEMDV-----NCC-KSNPC--ONDAICLDOI 470

QY 383 G--BCACRPGWSGLYCNET-----CSPGYGACQ--QIC 413
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 471 GGFHCICMPGEGYVFCQINSDDCASQPLCNKCIKDKINSFHCPCPKGSGSLCQVDDEC 530

QY 414 S---CQNCADCDVYTK--CTCAGPFGKIDC-----STPCPL-----GTYGINSSR 455
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 531 ASTPCKNGAKCTDPPNKTYCTECPFGSIGHCELDINECASSPCHYGVCRDGVASPTCDCR 590

QY 456 CG-----CKNDVACSPVDS--CTCAGHGVDCSR----- 485
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 591 PGYGRLCETININCLSQPCRNNGTCCODRENAVYICTCPKGTGVNCEINIDCKRRKPCDY 650

```

```

QY 486 -----CPSGTWGFGCNLTQ-----CLNGACNLTIDG-----TCTCAPGWRG 522
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 651 GKCLDKINGIECVCEPBGISGSMCNINIDDCALNPNCHNGTC--IDGVSPFCLCPDDEFRD 708

QY 523 EKC-----ELPCQDGTYYINC--AERC--DCSHADGCHP 552
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 709 ATCLSQHNECSSNPCIHGSCLDQINSRYCVCBAGMGNCDININECLSNPCVNGTCKD 768

QY 553 -TTGH--CCLGLGWSGVHDSVCABGRMP-----NCSL----- 583
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 769 MTSGYLTCRAGFGSPNCQNNINECASNPCLNQGSCIDVAGFKCNMCLPYTGEYCEVNL 828

QY 584 -PCY--CKNGASCSPDD-----GICE----- 601
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 829 APCSRPCPKNGVCRESDPFSFCNCPAGMGOTCEVDINECVARNPCTNGVCENLRG 888

QY 602 ---CAPGFRGTTQOR-----ICSPGYGHRCSQTCPOCV 632
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 889 FQCRPNPFGTALENDIDCEPNPCNSGVCOBVRNGFVCLAGFGERCAEDIDECV 948

QY 633 HSGPCHHITGLCCLPFTGALCENVCPSGRFGKNC--AGICT--CTNNGTCNPIDR 686
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 949 --SAPCRNGNCTDCVNSYT--CS--CPAGFSGINCEINPDTCESSCFNGGTC--VDG 999

QY 687 ---SCOCYPCWIGSDC-----SQP-----CPAHMGPNC---IH 714
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1000 ISSPSCVCLPFTGNYCHVDVNECDSPRCQNGSGCQDGYTKCKCPHGTGLNCSLYR 1059

QY 715 TCN---CHNGAFC--SAYDECKCTPGWTGLYCTQ----- 744
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1060 WCDSSPCPKNGSSCQMGQCAFQCCASGWTGYICDVPVSCEVAARQGVSAVILCRHAQ 1119

QY 745 -----RCPGFGYKDC--ALIQ-----QNGADC--DHISG--OCTCRTEGFRGRG 786
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1120 CYDAGNTHLCRQAGYTSYCOBOYDECPNCPONGATCTYLLGYSCECPVGHMGMS 1179

QY 787 QK-----CPSGTYGGRQICD-----CLNN 807
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1180 KEINECLSQPCQNGGTCTIDVNTYKSCSPRGTQGVHCHIDIDDSPSVDPLTGERCRNG 1239

QY 808 STC--DHITG--TCYCSPGKAKRCDQAGYIYGNLNSLSTALPADSY 854
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1240 GRCDVRVGGYGVCPAGFGERCE-----GDVNE--CLSDPCDPSGSY 1280

RESULT 9
NOTC_DROME STANDARD; PRT; 2703 AA.
AC P07207; P04154; Q97458; Q9M4T8;
DT 01-NOV-1986 (Rel. 03, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Neurogenic locus Notch protein precursor
GN N OR EG:140G11.1 OR EG:163A10.2 OR CG33936.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Oregon-R; TISSUE=Embryo;
RX MEDLINE=86079539; PubMed=393325;
RA Wharton K.A., Johansen K.M., Xu T., Artavanis-Tsakonas S.;
RT "Nucleotide sequence from the neurogenic locus notch implies a gene
RT product that shares homology with proteins containing EGF-like
RT repeats."
RL Cell 43:567-581(1985).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Canton-S; and Oregon-R; TISSUE=Embryo;
RX MEDLINE=87064624; PubMed=3097517;
RA Kidd S., Kelley M.R., Young M.W.;

```

RT "Sequence of the notch locus of *Drosophila melanogaster*: relationship
 RT of the encoded protein to mammalian clotting and growth factors.";
 RT Mol. Cell. Biol. 6:3094-3108(1986).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Berkeley;
 RX MEDLINE-20196006; PubMed-10731132;
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amaralides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Mortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.H.C., Blazer R.G., Chame M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Milos G.L.G.,
 RA Abiri J.F., Adayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borova D., Botchen M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burris K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durkin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jaisli M., Kaulush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai X.,
 RA Laasbo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin Z.,
 RA Liu X., Matzel B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merulov G., Milshine N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson K.A., Nelson K.A., Nixon K., Nusskern D.R., Pacleib J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reilner K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Klimos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang M., Zhao Q., Zhang L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RT Science 287:2185-2195(2000).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Oregon-R;
 RX MEDLINE-20196011; PubMed-10731137;
 RA Benos P.V., Gatt M.K., Ashburner M., Murphy L., Harris D.,
 RA Bartell B.G., Ferraz C., Vidal S., Brun C., Demallies J., Cadieu E.,
 RA Dreano S., Gloux S., Lelaure V., Motlier S., Galibert F., Borikova D.,
 RA Minna B., Kafatos F.C., Louis C., Siden-Klimos I., Bolshakov S.,
 RA Papagiannakis G., Spanos L., Cox S., Madueno E., de Pablos B.,
 RA Modellell J., Peter A., Schoettler P., Werner M., Murticott F.,
 RA Belhert N., Dowe G., Schaefer U., Jaekle H., Bucheton A.,
 RA Callister D.M., Campbell L.A., Darlamtsou A., Henderson N.S.,
 RA McMillan P.J., Salles C., Tait E.A., Valenti P., Saunders R.D.C.,
 RA Glover D.M.;
 RT "From sequence to chromosome: the tip of the X chromosome of *D.*
 RT *melanogaster*.";
 RT Science 287:2220-2222(2000).
 RN [5]
 RP SEQUENCE OF 2505-2611 FROM N.A.
 RX MEDLINE-8509329; PubMed-2981631;
 RA Wharton K.A., Yedvobnick B., Finerty V.G., Artavanis-Tsakonas S.;
 RT "Ops: a novel family of transcribed repeats shared by the Notch locus
 RT and other developmentally regulated loci in *D. melanogaster*.";
 RT Cell 40:55-62(1985).
 RN [6]
 RP SEQUENCE OF 1-8 FROM N.A.
 RX MEDLINE-87257846; PubMed-3037327;
 RA Kelley M.R., Kidd S., Berg R.L., Young M.W.;

RT "Restriction of P-element insertions at the Notch locus of *Drosophila*
 RT *melanogaster*.";
 RT Mol. Cell. Biol. 7:1545-1548(1987).
 RN [7]
 RP REVIEW.
 RA Harris W.A.;
 RT "Many cell types specified by Notch function.";
 RT Curr. Biol. 1:120-122(1991).
 CC -1- FUNCTION: Functions as a receptor for membrane-bound ligands Delta
 CC and Serrate to regulate cell-fate determination. Upon ligand
 CC activation through the released notch intracellular domain (NICD)
 CC it forms a transcriptional activator complex with Su(H)
 CC (suppressor of hairless) and activates genes of the enhancer of
 CC split locus. Essential for proper differentiation of ectoderm.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- SIMILARITY: BELONGS TO THE NOTCH FAMILY.
 CC -1- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
 CC -1- SIMILARITY: CONTAINS 6 ANK REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M6152; AAB59220.1; -
 DR EMBL: M6153; AAB59220.1; JOINED.
 DR EMBL: M6149; AAB59220.1; JOINED.
 DR EMBL: M6150; AAB59220.1; JOINED.
 DR EMBL: M6151; AAB59220.1; JOINED.
 DR EMBL: K03508; AAA28725.1; -
 DR EMBL: M13689; AAA28725.1; JOINED.
 DR EMBL: K03507; AAA28725.1; JOINED.
 DR EMBL: AE003426; AAF45848.2; -
 DR EMBL: AL035436; CAB37610.1; -
 DR EMBL: AL035395; CAB37610.1; JOINED.
 DR EMBL: M12175; AAA74496.1; -
 DR EMBL: M16025; AAA28726.1; -
 DR PIR: A24420; A24420.
 DR PIR: A24768; A24768.
 DR PIR: A05267; A05267.
 DR HSSP: P00740; IEDM.
 DR FLYBASE: FBgn0004647; N.
 DR InterPro: IPR002110; ANK.
 DR InterPro: IPR000152; Asx_hydroxyl.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR000742; EGF-2.
 DR InterPro: IPR001881; EGF_Ca.
 DR InterPro: IPR001438; EGF_II.
 DR InterPro: IPR002049; Lamln1_EGF.
 DR InterPro: IPR000800; Notch.
 DR Pfam: PR00008; EGF; 36.
 DR Pfam: PF00023; ank; 6.
 DR Pfam: PF00066; notch; 3.
 DR PRINTS: PR00010; EGFBLDOD.
 DR PRINTS: PR00011; EGF_LAMININ.
 DR PRINTS: PR01452; NOTCH.
 DR SMART: SM00248; ANK; 7.
 DR SMART: SM00181; EGF; 36.
 DR SMART: SM00179; EGF_Ca; 35.
 DR SMART: SM00004; NL; 2.
 DR PROSITE: PS50088; ANK_REPEAT; 5.
 DR PROSITE: PS50297; ANK_REPEAT_REGION; 1.
 DR PROSITE: PS00010; ASX_HYDROXYL; 22.
 DR PROSITE: PS00022; EGF_1; 34.
 DR PROSITE: PS01186; EGF_2; 28.
 DR PROSITE: PS01187; EGF_Ca; 21.
 KW Receptor: Transcription regulation: Activator: Repeat: ANK repeat;
 KW developmental protein: Neurogenesis: Repeat: ANK repeat;
 KW EGF-like domain: Transmembrane: Glycoprotein: Signal.


```

FT SIGNAL 1 44 POTENTIAL.
FT CHAIN 45 2703 NEUROGENIC LOCUS NOTCH PROTEIN.
FT DOMAIN 45 1745 EXTRACELLULAR (POTENTIAL).
FT TRANSHEM 1746 1766 POTENTIAL.
FT DOMAIN 1767 2703 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 58 95 EGF-LIKE 1.
FT DOMAIN 96 136 EGF-LIKE 2.
FT DOMAIN 139 176 EGF-LIKE 3.
FT DOMAIN 177 215 EGF-LIKE 4.
FT DOMAIN 217 253 EGF-LIKE 5.
FT DOMAIN 255 291 EGF-LIKE 6.
FT DOMAIN 293 329 EGF-LIKE 7.
FT DOMAIN 331 370 EGF-LIKE 8.
FT DOMAIN 372 408 EGF-LIKE 9.
FT DOMAIN 409 447 EGF-LIKE 10.
FT DOMAIN 449 486 EGF-LIKE 11.
FT DOMAIN 488 524 EGF-LIKE 12.
FT DOMAIN 526 562 EGF-LIKE 13.
FT DOMAIN 564 600 EGF-LIKE 14.
FT DOMAIN 602 637 EGF-LIKE 15.
FT DOMAIN 639 675 EGF-LIKE 16.
FT DOMAIN 677 713 EGF-LIKE 17.
FT DOMAIN 715 751 EGF-LIKE 18.
FT DOMAIN 753 789 EGF-LIKE 19.
FT DOMAIN 791 827 EGF-LIKE 20.
FT DOMAIN 829 865 EGF-LIKE 21.
FT DOMAIN 867 905 EGF-LIKE 22.
FT DOMAIN 907 944 EGF-LIKE 23.

```

Query Match 14.5%; Score 978.5; DB 1; Length 2703;
 Best Local Similarity 26.8%; Pred. No. 7.3e-50;
 Matches 290; Conservative 102; Mismatches 297; Indels 395; Gaps 70;

```

OY 7 SCL-----SFILLCHWIGTASPLNLED--FNVCSEHSESYSTVSESTPHPDQIYYISC 60
DB 502 SCLDDPGTFRCVCMGFTGTQCEIDIDECQSNPC-----LNDGTC 541
OY 61 TDILNMFECFTRHRSYRATYRHEGKTYRRKSQCCPFYSEGEHC-----VPHCADRCVHGR 117
DB 542 HDKINGFKCS-----CALGF--TGARKQINIDCQSPCCNR 576
OY 118 -----CIAPNTQCCEPFGNGTNC--SACDGDHMGPHCTSRQCKNALCNPING-ACH 168
DB 577 GICHDSIAGSECPGPGTGTSCIEINIDCSN-----PCHRGKCIDVNSFKCL 626
OY 169 CAAGRGM-----RCEDR-----CEGTYG-----NDCHQRCQ 196
DB 627 CDPGTGTGICQKQINECESNPQCFDGHQDRGSGYYCCQAGTSKNCSEVNVNECHSN-P 685
OY 197 CONGATC-DHYTG-ECRCPGTYGAFCEDLCPGKHGPOCEQRCPONGYVC-HHYTG-E 252
DB 686 CNGGATCIDGINSYKCCQCPVPGTGOHCE-----KNVDECIS-SPCANNGVCIDOVNIGK 738
OY 253 CSCPFGMNGTVC-----GQP-----CPGEGFNGNCS--OECC-- 282
DB 729 CECRPFDAHCLSDVDECAENPCVNEGRCEDEINEFICHCPRGYTGRCLEIDDECSSN 798
OY 283 -CHNGTC-DAATG-QCHCSFRTYGERCO--DECPVGTGYVLCAETQCYNAGKCY-HV 335
DB 799 PCQHGCTGYDKLNAFSCCMGTYGQCKETNIDDC-----VTNPGCNGGTCIDIKV 848
OY 336 SG-ACLCBAGFAGECEARLCPREGIYIKG-DKRCPLHENTHSHCPMSG--EACAKP 389
DB 849 NGKYCAVCVPTFTGRDCESKMDP-----CASNRC--KNEAKTPSSNFIIDFSCTKL 897
OY 390 GWSGLYCHNETCSPFGYBACQOICSCONGADQSDVTG--KCTCAPGFGGIDC----- 439
DB 898 GYTGRCDEDI-----DECSLSSPCRGASCLANVPSYRCICTGTYGGRGRCALMTDCA 951
OY 440 STPCP-----LGTYGINC-----SSRC-----GCKNDVCSPPVDGS--CTCK 474
DB 952 SFPCCNGGTCLDGIIDYDSCLCVDFGRHGHCTDINECLSPQONCATCSQYVNSTYCTCP 1011

```

```

OY 475 AGMHGVDCSIR-----CPSGTNGFGCN-----LTQC-----CLN 503
DB 1012 LGFSGINQNTDEDECTESSCLNGSGCINGINGINCYNCSLAGISGANCOYKILKKCSNPCLN 1071
OY 504 GGACNTLIDG--TCTCAPGMRGEC-----ELPCODGTGYGLNCAERCDSCSHADGCHPT 553
DB 1072 GATCHEQNNNEYTCHCPSGFTGKQKOCSEYVDMCGQSPCENG-----ATCSQMK--HQF 1120
OY 554 TGRCHCLPGRNGVHND-----SYCABEGRMGPNCISLPCTCKANGASCSP--DDGICECAPRG 608
DB 1121 S--CKCSAGMTGKGLDVTITSCODADRKGJSLROLCKNNG--TKRKYGNSHYVCSSQGYAG 1177
OY 609 TTCOR-----TCSPFYGHRSQTCPOCV-----HSSGPGH 639
DB 1178 SYCQKEIDECQSPONGSGTRDLGATECCQRRGFQONCELNIDDCAPRPNONGTICH 1237
OY 640 H--ITGLDCLPFTGALC-----NEVCPSGRFRGNKAGICTTNTGCTPRIDR--SCQC 690
DB 1238 DRYMNFSCSPPGTGTGIIICEINKDKCRG-----ACHNNQSC--IDRVGGEFCVC 1285
OY 691 YPGWIGSNC-----SQPCP-----PAHMGPCITCN----- 717
DB 1286 QPFGYAGCEEDINECLSNPCSNAGTLDVOLVNNHCNCRGHRGHRCEKRVDFCAOSP 1345
OY 718 CHNGAFCSAYDGECKCTPGMTGLYCTORCPGLFGYKDCALICQCONGACDHIS--GQC 774
DB 1346 CONGNCNIRQ-----SGHNCI--CNNGFYKNCEL-----SGDCDSNPRCVNKC 1389
OY 775 TCRTEGMRHCEQKCPSTGYGCR--QICD--CLNSTCHDITG--TCYCSPGWK 825
DB 1390 VVADEGFEYRCE--CPRSTLGEHCEIDITLDECSPPRCAQAGACEDLLDGYELCLPSKWK 1447
OY 826 ARCD 829
DB 1448 KRCD 1451

```

RESULT 10
 NTG3_MOUSE
 ID NTG3_MOUSE STANDARD; PRT; 2318 AA.
 AC Q61982;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Neurogenic locus notch homolog protein 3 precursor (Notch 3).
 GN NOTCH3.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ICR X Swiss Webster;
 RX MEDLINE=95001556; PubMed=7918097;
 RA Lardelli M., Dalstrand J., Lendahl U.;
 RT "The novel Notch homolog mouse Notch 3 lacks specific epidermal
 growth factor repeats and is expressed in proliferating
 neuroepithelium.";
 RL Mech. Dev. 46:123-136(1994).
 RN [2]
 RP POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS OF MET-1664.
 RX MEDLINE=21523956; PubMed=11518718;
 RA Saxena M.T., Schroeter E.H., Mumm J.S., Kopan R.;
 RT "Murine notch homologs (N1-4) undergo presenilin-dependent
 proteolysis.";
 RL J. Biol. Chem. 276:40268-40273(2001).
 RN [3]
 RP POST-TRANSLATIONAL PROCESSING.
 RX MEDLINE=21374376; PubMed=11459941;
 RA Mitutani T., Taniguchi Y., Aoki T., Hashimoto N., Honjo T.;
 RT "Conservation of the biochemical mechanisms of signal transduction
 among mammalian Notch family members.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9026-9031(2001).


```

FT DISULFID 380 389 BY SIMILARITY.
FT DISULFID 396 409 BY SIMILARITY.
FT DISULFID 403 418 BY SIMILARITY.
FT DISULFID 420 429 BY SIMILARITY.
FT DISULFID 436 447 BY SIMILARITY.
FT DISULFID 441 456 BY SIMILARITY.
FT DISULFID 458 467 BY SIMILARITY.
FT DISULFID 474 485 BY SIMILARITY.
FT DISULFID 479 494 BY SIMILARITY.
FT DISULFID 496 505 BY SIMILARITY.
FT DISULFID 512 523 BY SIMILARITY.
FT DISULFID 517 532 BY SIMILARITY.
FT DISULFID 534 543 BY SIMILARITY.
FT DISULFID 550 560 BY SIMILARITY.
FT DISULFID 555 569 BY SIMILARITY.
FT DISULFID 571 580 BY SIMILARITY.
FT DISULFID 587 598 BY SIMILARITY.
FT DISULFID 592 607 BY SIMILARITY.
FT DISULFID 609 618 BY SIMILARITY.
FT DISULFID 625 635 BY SIMILARITY.

```

Query Match 14.58; Score 977.5; DB 1; Length 2318;
 Best Local Similarity 25.0%; Pred. No. 7.2e-50;
 Matches 321; Conservative 78; Mismatches 309; Indels 575; Gaps 69;

```

OY 14 LILCHMIGTASPLNEDNVCSSHMSYSVTQESYHPHPDQIYYTSCDILNMFECTNRH 73
DB 30 LILLAGLAAAPCL-DSPCA-----NGRCHTHQ 59
OY 74 VSYRTAYRHEKTYRRKSQCCPGFYESGEMCPHCAKCVHGRCAIAPNTCO----- 125
DB 60 PLEA-----CICLPGMV--GERC--QLEDPCHSPPAGAGVQSSVYAGTA 103
OY 136 -----CEGWMGTNCSSACDGDHMGPHCTSRCCCKNALCNPTG-----ACHCAAGFRG 175
DB 104 RESCRLKGFQGPDCSOP-----DPCVSR-PCVHGAPCS--VGPDRFACACPPGYOG 153
OY 176 WRCE-----DRCEGTGNGDNHOCOCONGATCDHTG--ECRCPPRYTAFCD--LPP 227
DB 154 OSCOSDIDECNRGT-----TCRHGTCLNTPGSRCCPLIYTLILCENPVVPCAP 204
OY 228 GKHGPCBQRCPCQNGVCH--VTGBCSCPSPGMGTGCG--OPCEGFRGKNCQEC 281
DB 205 S-----PCRNNGTGRROSSDVTYDCACLPFGEGNCEVNVDDCGHR----- 245
OY 252 QCHNGTC--DAATGQCCHSPRYTGERCO--DECPVGTGYVLCAETCCVAGGCKYHS 336
DB 246 -CLNGTCTVDGVNTYNNCCCPREMTGOFCTEDVDECOL-----QPNACHNGGTCFNL 296
OY 337 G--ACICEAGFAGERCERL-----CPEGLYGKIC--DKRC 368
DB 257 GGHSCVCVNGWTGESCSONIDDCATAVCFHATCDHRAVSFACACPMKRTGLLCHLDAC 356
OY 359 ---PCHLENTSHCHPMSE--CACRPGWSGLYCNE----- 398
DB 357 VSNPCHDAICDTPNVSRAICTCPFGTGACDQDVDECISGANPCEHLGRCVNTQSGF 416
OY 359 -----TCSPPFYBACQ-----QI 412
DB 417 LCQGGRTYGPCECTDVNECLSGPCRNATCLDRIGQFTCIOMAFETGYCEVDIDECOS 476
OY 413 CSCQAGADC-DSVTG-KCTCAAPFKGIDC-----STPCPLGTG-----GNCSSNC 456
DB 477 SPCVAGVCKDRVNGFSCTCBGFSGSMQOLDVDECASTPCPNNGAKCYDQDPGYEC--RC 534
OY 457 G-----C-KNDVACSP-----VDG--SCTCRAGMHGVDCS----- 483
DB 535 AEGFGTLCERNVDCSPDCHHGRCVDIASFCACAPGYGICGCEQVDECRSQPCRY 594
OY 484 -----IRCPSTGWGSCNLTG--QCLANG-----GACNTLDG-----TCTCAPGRNG 522
DB 595 GSKCLDLVKKYLCRCPPTGTGNCVNIIDDCASNPCTPVGCR--DGINKRYDCVCPGFTG 652

```

```

OY 523 EXCEL-----PCODGTGILNC-----AERDCSHADGCH 551
DB 653 PLCNWEINDECASSPCBEGSGCYDGBNGFRHCLCPGSLPPLCLPANHPCAHKPCSHG-VCH 711
OY 552 PTTG--HCRCLPGWSGVHCDVYACABGRWPNCSLPCYCKNGASCSPDDGI--CECARGF 606
DB 712 DAPGFRVCERCEGWSGPCOSGLA-----PDACESQPOAGGTCT-SBIGFRCTCARGF 765
OY 607 KCTTCQRI-----GSPGYGHRCSQTCPCQCHSS-----G 636
DB 766 QGHQCEVSPCTPSLCEHGHCESDPDRILTVCSCPGMOGPRCQDQDVDEACASPCGPHG 825
OY 637 PCNHTTG--LDCCLPGFALCNE-----YCPSGFRKNC 669
DB 826 TCTNLPGRFCLCHRGTYGPPCDQDIDDCDPNPLCHGSCQDGVGSFSCSLDGFAGPRC 885
OY 670 A-----GICP-----CTNNGTCNPIDR 686
DB 886 ARVDDECLSPGCPPTCTDHTVASFACAPPGYGFHCEIDLPDSCSPSCFNCGTC--VDG 943
OY 687 -----SCCYPGNIGSDC-----SQP-----CPRAHMGPCNLIHTCN----- 717
DB 944 VSSFECILCRPGYGTGHCOYEADPCFSRPLHGLICNPTHPGFEC--TCREGFTGSCQNP 1001
OY 718 -----CHNGAPCSAYDGECKCPRGWTG-----LYCQR----- 745
DB 1002 VDMCSQAPCQNGRCVGTGANTCICPPGWSGRCLDIQSLPTEAANAOMGVREQLCQEGCK 1061
OY 746 -----CPLGYGRKC-----ALICQONGADCDHISG--QCTCRTGFMGRCE 786
DB 1062 CIDKRSXYCVCPBEGRTSHCHEYDPTAQCQHGICRGVWGVCBPCRGVAGDSCE 1121
OY 787 Q-----KCPSTGYGIC--ROICD-----CLNNS 808
DB 1122 DNIDECASQPCQNGSCSIDLVARYLCSPPGTGLVLCIEINEDCDLGPSLDSGVQCLHNG 1181
OY 809 TCDHITG--TCYCSPPWKGARCD 829
DB 1182 TCVDLVGFRNCPPGYTGHLCE 1204

```

RESULT 11
 NTC3_HUMAN STANDARD: PRT: 2321 AA.
 AC Q9UM47; Q9U618; Q9UPL3; Q9UEB3;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DE 15-JUN-2002 (Rel. 41, Last annotation update)
 GN Neurogenic locus notch homolog protein 3 precursor (Notch 3).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP MEDLINE=97032728; PubMed=8878478;
 RA Joutel A., Corpechot C., Ducros A., Vahedi K., Chabriat H., Mouton P.,
 Alamiotich S., Domenga V., Cecillon M., Marechal E., MacLarek J.,
 Vayssiere C., Cruaud C., Cabanis E.-A., Ruchoux M.M., Weissbach J.,
 Bach J.-F., Bousset M.-G., Tourtelier-Lasserre E.,
 RT "Notch3 mutations in CADASIL, a hereditary adult-onset condition
 causing stroke and dementia".
 RL Nature 385:707-710(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Guenel M., Artavanis-Tsakonas S.;
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Lamerdin J.E., McCreedy P.M., Skowronski E., Adamson A.W.,
 RA Burkhardt-Schultz K., Gordon L., Ramirez M., Stillwagen S.,
 RA Phan H., Velasco N., Garnes J., Dangnanan L., Poundstone P.,

QY	107	PHCAD-----	-ACVHRCIAPRTCCCEPMEGTSTSSACDGHMPCHTSCQCNGA	158
Db	42	PCCLDSCPCANGSGCTHQDPREAAACICLCPWVHEROLE-DPCHSFP-CAGRIYCOSSV	99	
QY	159	LCNPITGACHCACAGFRGRCE--DRC-----	EQGTG-----	188
Db	100	VAGVARESCRLRFRFGDDCSLPRPCESSPACAGAPCSVSGDGRYKACACPRGYGRNCRS	159	
QY	189	---NPDHRCOCOMATGDHVTG---ECCRCPPYTAFCED---LCRPGKHGPCCQRPCQ	241	
Db	160	DIDECRAGASCRHGCTGCTNTPGSHHCLCPILGTYLLDENLTPVCPAPF-----PCR	209	
QY	242	NGCYCHH---VTGECSCPSGMMGTVCG---QPCBGRFKNCSQECCHNGGTC--DAAT	293	
Db	210	NGGTCROSSDITYACACLPGEFGNCEVNNDDCGHR-----CLMNGTCVDGVNT	255	
QY	294	GQCHSPGYTERDO---DECPCVTGYLAETQCYNNGGCKYHVS--ACLCAPAGE	348	
Db	260	YNCCCPREMTQFCTEDYDECQL-----QPNACHNGGTCFNLHGSHSCVYANGTGE	311	
QY	349	RCEARL-----CPRELYGKIC--DKRC---PCHLENTSCH	379	
Db	312	SCSONIDCATCAVCFHGATCDHRAVSFACAPMGKTGLLCHLDACVSNPCHEDAICDTN	371	
QY	380	PMSEB--CACKRGMSSGLYCNF-----	398	
Db	372	FVSGRAITCPRPFTGGACDDVDCESTIGANPCENHLRCVNTGSGFLCQCGRGTYGRCE	431	
QY	399	-----TCSPFYEACQ-----QICSCONGADC--DSYT	425	
Db	432	TDVNECLSGPCRNQATCLDRICQFTCIOMAGFTGTFCEVVIDEBOQSSPCVYNGVCKRNVN	491	
QY	426	G-KCTCAGAFGGIDC-----STCPRLGYT-----GINCSSRCG-----C-RNDA	462	
Db	492	GFSCTCPBGSFGSCTOLDVDECASTPCRNKAKCYDQPDGYEC--RCABEGEGTLCERNVD	549	
QY	463	VCSP-----VYG-----SCTCAGMHGVDCS-----IR	485	
Db	550	DCSPDPCRHGKHCVDGINSFSCACAPGTTGTIRCSQVDECHSQPCRRGGCKLIDLVDKYLR	609	
QY	486	CPSPSTGFGCNLTG-QCLNG---GACNTLDG---TCTCAPGRGKECEL-----	527	
Db	610	CPPTGTGNCVENVINDDCASNPCTFGVGR--DGIRNYDCVCPGPTGTLCAVNEINECASSP	667	
QY	528	-----PCDDGYTGLNC-----AERDCSHADGCHPTTG--HCRCLPGW	563	
Db	668	CGBGSGCVDGNGHNCIOLCPRESLPRILCANHPHCAKPCSHG--VCHDAPGGFOCYDDPGW	726	

RT "Human notch4 gene

RA	Mt
----	----

RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 [4]
 RP IDENTIFICATION OF LIGANDS.
 RA MEDLINE-99180765; PubMed-10079256;
 RA Gray G.E., Mann R.S., Mitsiadis E., Henrique D., Carcangiu M.-L.,
 RA Banks A., Leiman J., Ward D., Ish-Horowitz D., Arlavanis-Tsakonas S.,
 RT Human ligands of the Notch receptor.
 RT Am. J. Pathol. 154:785-794(1999).
 CC -1- FUNCTION: Functions as a receptor for membrane-bound ligands
 CC Jagged1, Jagged2 and Delta to regulate cell-fate determination.
 CC upon ligand activation through the released notch intracellular
 CC domain (NICD). It forms a transcriptional activator complex with
 CC RBP-J kappa and activates genes of the enhancer of split locus.
 CC Affects the implementation of differentiation, proliferation and
 CC apoptotic programs. May regulate branching morphogenesis in the
 CC developing vascular system (By similarity).
 CC -1- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
 CC terminal fragment N(EC) which are probably linked by disulfide
 CC bonds (By similarity).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Following
 CC proteolytical processing NICD is translocated to the nucleus.
 CC -1- ALTERNATIVE PRODUCTS: 3 isoforms: 1 (shown here), 2 and 3; may be
 CC produced by alternative splicing.
 CC -1- TISSUE SPECIFICITY: Highly expressed in the heart, moderately in
 CC the lung and placenta and at low levels in the liver, skeletal
 CC muscle, kidney, pancreas, spleen, lymph node, thymus, bone marrow
 CC and fetal liver. No expression was seen in adult brain or
 CC peripheral blood leukocytes.
 CC -1- PTM: Synthesized in the endoplasmic reticulum as an inactive form
 CC which is proteolytically cleaved by a furin-like convertase in the
 CC trans-golgi network before it reaches the plasma membrane to yield
 CC an active, ligand-accessible form. Cleavage results in a C-
 CC terminal fragment N(TM) and a N-terminal fragment N(EC). Following
 CC ligand binding, it is cleaved by TNF-alpha converting enzyme
 CC (TACE) to yield a membrane-associated intermediate fragment called
 CC notch extracellular truncation (NEXT). This fragment is then
 CC cleaved by presentin dependent gamma-secretase to release a
 CC notch-derived peptide containing the intracellular domain (NICD)
 CC from the membrane (By similarity).
 CC -1- PTM: Phosphorylated (By similarity).
 CC -1- POLYMORPHISM: The poly-Leu region of NOTCH4 (In the signal
 CC peptide) is polymorphic and the number of Leu varies in the
 CC population (from 6 to 12).
 CC -1- SIMILARITY: BELONGS TO THE NOTCH FAMILY.
 CC -1- SIMILARITY: CONTAINS 28 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
 CC -1- SIMILARITY: CONTAINS 5 ANK REPEATS.
 CC -1- CAUTION: Ref.1 sequence differs from that shown due to frameshifts
 CC in position 1438 to 1463.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL; D63395; BAA09708.1; ALT_FRAME.
 CC EMBL; D86566; BAA13116.1; -
 CC EMBL; U95299; AAC32288.1; -
 CC EMBL; U89335; AAC33097.1; -
 CC EMBL; AB023961; BAB20317.1; -
 CC EMBL; AB024520; BAA88951.1; -
 CC EMBL; AB024578; BAA88952.1; -
 CC HSSP; P08709; IBP9.
 CC Genev; HGNC:7884; NOTCH4.
 CC MIM; 164951; -
 CC InterPro; IPR002110; ANK.
 CC InterPro; IPR000152; Asx_hydroxyl.
 CC InterPro; IPR000561; EGF-like.
 CC InterPro; IPR000742; EGF_2.
 CC InterPro; IPR001881; EGF_Ca.

DR InterPro; IPR001438; EGF-II.
 DR InterPro; IPR000800; Notch.
 DR Pfam; PF00008; EGF; 26.
 DR Pfam; PF00023; ank; 6.
 DR PRINTS; PR00066; notch_2.
 DR PRINTS; PR00010; EGFBLD.
 DR PRINTS; PR00011; EGFFLAMIN.
 DR PRINTS; PR00012; ENTPEI.
 DR SMART; SM00248; ANK; 5.
 DR SMART; SM00179; EGF_CA; 11.
 DR SMART; SM00001; EGF_Like; 15.
 DR SMART; SM00004; NL; 2.
 DR PROSITE; PS50088; ANK_REPEAT; 5.
 DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
 DR PROSITE; PS00010; ASX_HYDROXYL; 11.
 DR PROSITE; PS00022; EGF_1; 28.
 DR PROSITE; PS01186; EGF_2; 21.
 DR PROSITE; PS01187; EGF_CA; 9.
 KW Receptor; Transcription regulation; Activator; Differentiation;
 KW Developmental protein; Repeat; ANK repeat; EGF-like domain;
 KW Transmembrane; Glycoprotein; Signal; Phosphorylation; Polymorphism;
 KW Triplet repeat expansion; Alternative splicing.
 FT SIGNAL 1 23
 FT CHAIN 24 2003
 FT CHAIN 1432 2003
 FT CHAIN 1467 2003
 FT DOMAIN 24 1447
 FT TRANSMEM 1448 1468
 FT DOMAIN 1469 2003
 FT DOMAIN 24 63
 FT DOMAIN 64 115
 FT DOMAIN 118 155
 FT DOMAIN 156 192
 FT DOMAIN 194 232
 FT DOMAIN 234 274
 FT DOMAIN 276 312
 FT DOMAIN 314 353
 FT DOMAIN 355 391
 FT DOMAIN 392 430
 FT DOMAIN 432 473
 FT DOMAIN 475 511
 FT DOMAIN 513 549
 FT DOMAIN 551 587
 FT DOMAIN 589 625
 FT DOMAIN 626 659
 FT DOMAIN 661 689
 FT DOMAIN 691 727
 FT DOMAIN 729 765
 FT DOMAIN 767 803
 FT DOMAIN 806 842
 FT DOMAIN 844 880
 FT DOMAIN 882 928
 FT DOMAIN 930 966
 FT DOMAIN 968 1004
 FT DOMAIN 1006 1044
 FT DOMAIN 1046 1085
 FT DOMAIN 1087 1126
 FT DOMAIN 1130 1171
 FT DOMAIN 1173 1212
 FT REPEAT 1213 1246
 FT REPEAT 1247 1286
 FT REPEAT 1287 1326
 FT REPEAT 1327 1365
 FT REPEAT 1366 1405
 FT REPEAT 1406 1445
 FT REPEAT 1446 1485
 FT REPEAT 1486 1525
 FT REPEAT 1526 1565
 FT REPEAT 1566 1605
 FT REPEAT 1606 1645
 FT REPEAT 1646 1685
 FT REPEAT 1686 1725
 FT REPEAT 1726 1765
 FT REPEAT 1766 1805
 FT DISULFID 28 41
 FT DISULFID 35 51
 FT DISULFID 53 62
 FT DISULFID 68 80
 FT CHAIN 1467 2003
 FT CHAIN 1432 2003
 FT CHAIN 1467 2003
 FT DOMAIN 24 1447
 FT TRANSMEM 1448 1468
 FT DOMAIN 1469 2003
 FT DOMAIN 24 63
 FT DOMAIN 64 115
 FT DOMAIN 118 155
 FT DOMAIN 156 192
 FT DOMAIN 194 232
 FT DOMAIN 234 274
 FT DOMAIN 276 312
 FT DOMAIN 314 353
 FT DOMAIN 355 391
 FT DOMAIN 392 430
 FT DOMAIN 432 473
 FT DOMAIN 475 511
 FT DOMAIN 513 549
 FT DOMAIN 551 587
 FT DOMAIN 589 625
 FT DOMAIN 626 659
 FT DOMAIN 661 689
 FT DOMAIN 691 727
 FT DOMAIN 729 765
 FT DOMAIN 767 803
 FT DOMAIN 806 842
 FT DOMAIN 844 880
 FT DOMAIN 882 928
 FT DOMAIN 930 966
 FT DOMAIN 968 1004
 FT DOMAIN 1006 1044
 FT DOMAIN 1046 1085
 FT DOMAIN 1087 1126
 FT DOMAIN 1130 1171
 FT DOMAIN 1173 1212
 FT REPEAT 1213 1246
 FT REPEAT 1247 1286
 FT REPEAT 1287 1326
 FT REPEAT 1327 1365
 FT REPEAT 1366 1405
 FT REPEAT 1406 1445
 FT REPEAT 1446 1485
 FT REPEAT 1486 1525
 FT REPEAT 1526 1565
 FT REPEAT 1566 1605
 FT REPEAT 1606 1645
 FT REPEAT 1646 1685
 FT REPEAT 1686 1725
 FT REPEAT 1726 1765
 FT REPEAT 1766 1805
 FT DISULFID 28 41
 FT DISULFID 35 51
 FT DISULFID 53 62
 FT DISULFID 68 80
 FT CHAIN 1467 2003
 FT CHAIN 1432 2003
 FT CHAIN 1467 2003
 FT DOMAIN 24 1447
 FT TRANSMEM 1448 1468
 FT DOMAIN 1469 2003
 FT DOMAIN 24 63
 FT DOMAIN 64 115
 FT DOMAIN 118 155
 FT DOMAIN 156 192
 FT DOMAIN 194 232
 FT DOMAIN 234 274
 FT DOMAIN 276 312
 FT DOMAIN 314 353
 FT DOMAIN 355 391
 FT DOMAIN 392 430
 FT DOMAIN 432 473
 FT DOMAIN 475 511
 FT DOMAIN 513 549
 FT DOMAIN 551 587
 FT DOMAIN 589 625
 FT DOMAIN 626 659
 FT DOMAIN 661 689
 FT DOMAIN 691 727
 FT DOMAIN 729 765
 FT DOMAIN 767 803
 FT DOMAIN 806 842
 FT DOMAIN 844 880
 FT DOMAIN 882 928
 FT DOMAIN 930 966
 FT DOMAIN 968 1004
 FT DOMAIN 1006 1044
 FT DOMAIN 1046 1085
 FT DOMAIN 1087 1126
 FT DOMAIN 1130 1171
 FT DOMAIN 1173 1212
 FT REPEAT 1213 1246
 FT REPEAT 1247 1286
 FT REPEAT 1287 1326
 FT REPEAT 1327 1365
 FT REPEAT 1366 1405
 FT REPEAT 1406 1445
 FT REPEAT 1446 1485
 FT REPEAT 1486 1525
 FT REPEAT 1526 1565
 FT REPEAT 1566 1605
 FT REPEAT 1606 1645
 FT REPEAT 1646 1685
 FT REPEAT 1686 1725
 FT REPEAT 1726 1765
 FT REPEAT 1766 1805
 FT DISULFID 28 41
 FT DISULFID 35 51
 FT DISULFID 53 62
 FT DISULFID 68 80
 FT CHAIN 1467 2003
 FT CHAIN 1432 2003
 FT CHAIN 1467 2003
 FT DOMAIN 24 1447
 FT TRANSMEM 1448 1468
 FT DOMAIN 1469 2003
 FT DOMAIN 24 63
 FT DOMAIN 64 115
 FT DOMAIN 118 155
 FT DOMAIN 156 192
 FT DOMAIN 194 232
 FT DOMAIN 234 274
 FT DOMAIN 276 312
 FT DOMAIN 314 353
 FT DOMAIN 355 391
 FT DOMAIN 392 430
 FT DOMAIN 432 473
 FT DOMAIN 475 511
 FT DOMAIN 513 549
 FT DOMAIN 551 587
 FT DOMAIN 589 625
 FT DOMAIN 626 659
 FT DOMAIN 661 689
 FT DOMAIN 691 727
 FT DOMAIN 729 765
 FT DOMAIN 767 803
 FT DOMAIN 806 842
 FT DOMAIN 844 880
 FT DOMAIN 882 928
 FT DOMAIN 930 966
 FT DOMAIN 968 1004
 FT DOMAIN 1006 1044
 FT DOMAIN 1046 1085
 FT DOMAIN 1087 1126
 FT DOMAIN 1130 1171
 FT DOMAIN 1173 1212
 FT REPEAT 1213 1246
 FT REPEAT 1247 1286
 FT REPEAT 1287 1326
 FT REPEAT 1327 1365
 FT REPEAT 1366 1405
 FT REPEAT 1406 1445
 FT REPEAT 1446 1485
 FT REPEAT 1486 1525
 FT REPEAT 1526 1565
 FT REPEAT 1566 1605
 FT REPEAT 1606 1645
 FT REPEAT 1646 1685
 FT REPEAT 1686 1725
 FT REPEAT 1726 1765
 FT REPEAT 1766 1805
 FT DISULFID 28 41
 FT DISULFID 35 51
 FT DISULFID 53 62
 FT DISULFID 68 80
 FT CHAIN 1467 2003
 FT CHAIN 1432 2003
 FT CHAIN 1467 2003
 FT DOMAIN 24 1447
 FT TRANSMEM 1448 1468
 FT DOMAIN 1469 2003
 FT DOMAIN 24 63
 FT DOMAIN 64 115
 FT DOMAIN 118 155
 FT DOMAIN 156 192
 FT DOMAIN 194 232
 FT DOMAIN 234 274
 FT DOMAIN 276 312
 FT DOMAIN 314 353
 FT DOMAIN 355 391
 FT DOMAIN 392 430
 FT DOMAIN 432 473
 FT DOMAIN 475 511
 FT DOMAIN 513 549
 FT DOMAIN 551 587
 FT DOMAIN 589 625
 FT DOMAIN 626 659
 FT DOMAIN 661 689
 FT DOMAIN 691 727
 FT DOMAIN 729 765
 FT DOMAIN 767 803
 FT DOMAIN 806 842
 FT DOMAIN 844 880
 FT DOMAIN 882 928
 FT DOMAIN 930 966
 FT DOMAIN 968 1004
 FT DOMAIN 1006 1044
 FT DOMAIN 1046 1085
 FT DOMAIN 1087 1126
 FT DOMAIN 1130 1171
 FT DOMAIN 1173 1212
 FT REPEAT 1213 1246
 FT REPEAT 1247 1286
 FT REPEAT 1287 1326
 FT REPEAT 1327 1365
 FT REPEAT 1366 1405
 FT REPEAT 1406 1445
 FT REPEAT 1446 1485
 FT REPEAT 1486 1525
 FT REPEAT 1526 1565
 FT REPEAT 1566 1605
 FT REPEAT 1606 1645
 FT REPEAT 1646 1685
 FT REPEAT 1686 1725
 FT REPEAT 1726 1765
 FT REPEAT 1766 1805
 FT DISULFID 28 41
 FT DISULFID 35 51
 FT DISULFID 53 62
 FT DISULFID 68 80
 FT CHAIN 1467 2003
 FT CHAIN 1432 2003
 FT CHAIN 1467 2003
 FT DOMAIN 24 1447
 FT TRANSMEM 1448 1468
 FT DOMAIN 1469 2003
 FT DOMAIN 24 63
 FT DOMAIN 64 115
 FT DOMAIN 118 155
 FT DOMAIN 156 192
 FT DOMAIN 194 232
 FT DOMAIN 234 274
 FT DOMAIN 276 312
 FT DOMAIN 314 353
 FT DOMAIN 355 391
 FT DOMAIN 392 430
 FT DOMAIN 432 473
 FT DOMAIN 475 511
 FT DOMAIN 513 549
 FT DOMAIN 551 587
 FT DOMAIN 589 625
 FT DOMAIN 626 659
 FT DOMAIN 661 689
 FT DOMAIN 691 727
 FT DOMAIN 729 765
 FT DOMAIN 767 803
 FT DOMAIN 806 842
 FT DOMAIN 844 880
 FT DOMAIN 882 928
 FT DOMAIN 930 966
 FT DOMAIN 968 1004
 FT DOMAIN 1006 1044
 FT DOMAIN 1046 1085
 FT DOMAIN 1087 1126
 FT DOMAIN 1130 1171
 FT DOMAIN 1173 1212
 FT REPEAT 1213 1246
 FT REPEAT 1247 1286
 FT REPEAT 1287 1326
 FT REPEAT 1327 1365
 FT REPEAT 1366 1405
 FT REPEAT 1406 1445
 FT REPEAT 1446 1485
 FT REPEAT 1486 1525
 FT REPEAT 1526 1565
 FT REPEAT 1566 1605
 FT REPEAT 1606 1645
 FT REPEAT 1646 1685
 FT REPEAT 1686 1725
 FT REPEAT 1726 1765
 FT REPEAT 1766 1805
 FT DISULFID 28 41
 FT DISULFID 35 51
 FT DISULFID 53 62
 FT DISULFID 68 80
 FT CHAIN 1467 2003
 FT CHAIN 1432 2003
 FT CHAIN 1467 2003
 FT DOMAIN 24 1447
 FT TRANSMEM 1448 1468
 FT DOMAIN 1469 2003
 FT DOMAIN 24 63
 FT DOMAIN 64 115
 FT DOMAIN 118 155
 FT DOMAIN 156 192
 FT DOMAIN 194 232
 FT DOMAIN 234 274
 FT DOMAIN 276 312
 FT DOMAIN 314 353
 FT DOMAIN 355 391
 FT DOMAIN 392 430
 FT DOMAIN 432 473
 FT DOMAIN 475 511
 FT DOMAIN 513 549
 FT DOMAIN 551 587
 FT DOMAIN 589 625
 FT DOMAIN 626 659
 FT DOMAIN 661 689
 FT DOMAIN 691 727
 FT DOMAIN 729 765
 FT DOMAIN 767 803
 FT DOMAIN 806 842
 FT DOMAIN 844 880
 FT DOMAIN 882 928
 FT DOMAIN 930 966
 FT DOMAIN 968 1004
 FT DOMAIN 1006 1044
 FT DOMAIN 1046 1085
 FT DOMAIN 1087 1126
 FT DOMAIN 1130 1171
 FT DOMAIN 1173 1212
 FT REPEAT 1213 1246
 FT REPEAT 1247 1286
 FT REPEAT 1287 1326
 FT REPEAT 1327 1365
 FT REPEAT 1366 1405
 FT REPEAT 1406 1445
 FT REPEAT 1446 1485
 FT REPEAT 1486 1525
 FT REPEAT 1526 1565
 FT REPEAT 1566 1605
 FT REPEAT 1606 1645
 FT REPEAT 1646 1685
 FT REPEAT 1686 1725
 FT REPEAT 1726 1765
 FT REPEAT 1766 1805
 FT DISULFID 28 41
 FT DISULFID 35 51
 FT DISULFID 53 62
 FT DISULFID 68 80
 FT CHAIN 1467 2003
 FT CHAIN 1432 2003
 FT CHAIN 1467 2003
 FT DOMAIN 24 1447
 FT TRANSMEM 1448 1468
 FT DOMAIN 1469 2003
 FT DOMAIN 24 63
 FT DOMAIN 64 115
 FT DOMAIN 118 155
 FT DOMAIN 156 192
 FT DOMAIN 194 232
 FT DOMAIN 234 274
 FT DOMAIN 276 312
 FT DOMAIN 314 353
 FT DOMAIN 355 391
 FT DOMAIN 392 430
 FT DOMAIN 432 473
 FT DOMAIN 475 511
 FT DOMAIN 513 549
 FT DOMAIN 551 587
 FT DOMAIN 589 625
 FT DOMAIN 626 659
 FT DOMAIN 661 689
 FT DOMAIN 691 727
 FT DOMAIN 729 765
 FT DOMAIN 767 803
 FT DOMAIN 806 842
 FT DOMAIN 844 880
 FT DOMAIN 882 928
 FT DOMAIN 930 966
 FT DOMAIN 968 1004
 FT DOMAIN 1006 1044
 FT DOMAIN 1046 1085
 FT DOMAIN 1087 1126
 FT DOMAIN 1130 1171
 FT DOMAIN 1173 1212
 FT REPEAT 1213 1246
 FT REPEAT 1247 1286
 FT REPEAT 1287 1326
 FT REPEAT 1327 1365
 FT REPEAT 1366 1405
 FT REPEAT 1406 1445
 FT REPEAT 1446 1485
 FT REPEAT 1486 1525
 FT REPEAT 1526 1565
 FT REPEAT 1566 1605
 FT REPEAT 1606 1645
 FT REPEAT 1646 1685
 FT REPEAT 1686 1725
 FT REPEAT 1726 1765
 FT REPEAT 1766 1805
 FT DISULFID 28 41
 FT DISULFID 35 51
 FT DISULFID 53 62
 FT DISULFID 68 80
 FT CHAIN 1467 2003
 FT CHAIN 1432 2003
 FT CHAIN 1467 2003
 FT DOMAIN 24 1447
 FT TRANSMEM 1448 1468
 FT DOMAIN 1469 2003
 FT DOMAIN 24 63
 FT DOMAIN 64 115
 FT DOMAIN 118 155
 FT DOMAIN 156 192
 FT DOMAIN 194 232
 FT DOMAIN 234 274
 FT DOMAIN 276 312
 FT DOMAIN 314 353
 FT DOMAIN 355 391
 FT DOMAIN 392 430
 FT DOMAIN 432 473
 FT DOMAIN 475 511
 FT DOMAIN 513 549
 FT DOMAIN 551 587
 FT DOMAIN 589 625
 FT DOMAIN 626 659
 FT DOMAIN 661 689
 FT DOMAIN 691 727
 FT DOMAIN 729 765
 FT DOMAIN 767 803
 FT DOMAIN 806 842
 FT DOMAIN 844 880
 FT DOMAIN 882 928
 FT DOMAIN 930 966
 FT DOMAIN 968 1004
 FT DOMAIN 1006 1044
 FT DOMAIN 1046 1085
 FT DOMAIN 1087 1126
 FT DOMAIN 1130 1171
 FT DOMAIN 1173 1212
 FT REPEAT 1213 1246
 FT REPEAT 1247 1286
 FT REPEAT 1287 1326
 FT REPEAT 1327 1365
 FT REPEAT 1366 1405
 FT REPEAT 1406 1445
 FT REPEAT 1446 1485
 FT REPEAT 1486 1525
 FT REPEAT 1526 1565
 FT REPEAT 1566 1605
 FT REPEAT 1606 1645
 FT REPEAT 1646 1685
 FT REPEAT 1686 1725
 FT REPEAT 1726 1765
 FT REPEAT 1766 1805
 FT DISULFID 28 41
 FT DISULFID 35 51
 FT DISULFID 53 62
 FT DISULFID 68 80
 FT CHAIN 1467 2003
 FT CHAIN 1432 2003
 FT CHAIN 1467 2003
 FT DOMAIN 24 1447
 FT TRANSMEM 1448 1468
 FT DOMAIN 1469 2003
 FT DOMAIN 24 63
 FT DOMAIN 64 115
 FT DOMAIN 118 155
 FT DOMAIN 156 192
 FT DOMAIN 194 232
 FT DOMAIN 234 274
 FT DOMAIN 276 312
 FT DOMAIN 314 353
 FT DOMAIN 355 391
 FT DOMAIN 392 430
 FT DOMAIN 432 473
 FT DOMAIN 475 511
 FT DOMAIN 513 549
 FT DOMAIN 551 587
 FT DOMAIN 589 625
 FT DOMAIN 626 659
 FT DOMAIN 661 689
 FT DOMAIN 691 727
 FT DOMAIN 729 765
 FT DOMAIN 767 803
 FT DOMAIN 806 842
 FT DOMAIN 844 880
 FT DOMAIN 882 928
 FT DOMAIN 930 966
 FT DOMAIN 968 1004
 FT DOMAIN 1006 1044
 FT DOMAIN 1046 1085
 FT DOMAIN 1087 1126
 FT DOMAIN 1130 1171
 FT DOMAIN 1173 1212
 FT REPEAT 1213 1246
 FT REPEAT 1247 1286
 FT REPEAT 1287 1326
 FT REPEAT 1327 1365
 FT REPEAT 1366 1405
 FT REPEAT 1406 1445
 FT REPEAT 1446 1485
 FT REPEAT 1486 1525
 FT REPEAT 1526 1565
 FT REPEAT 1566 1605
 FT REPEAT 1606 1645
 FT REPEAT 1646 1685
 FT REPEAT 1686 1725
 FT REPEAT 1726 1765
 FT REPEAT 1766 1805
 FT DISULFID 28 41
 FT DISULFID 35 51
 FT DISULFID 53 62
 FT DISULFID 68 80
 FT CHAIN 1467 2003
 FT CHAIN 1432 2003
 FT CHAIN 1467 2003
 FT DOMAIN 24 1447
 FT TRANSMEM 1448 1468
 FT DOMAIN 1469 2003
 FT DOMAIN 24 63
 FT DOMAIN 64 115
 FT DOMAIN 118 155
 FT DOMAIN 156 192
 FT DOMAIN 194 232
 FT DOMAIN 234 274
 FT DOMAIN 276 312
 FT DOMAIN 314 353
 FT DOMAIN 355 391
 FT DOMAIN 392 430
 FT DOMAIN 432 473
 FT DOMAIN 475 511
 FT DOMAIN 513 549
 FT DOMAIN 551 587
 FT DOMAIN 589 625
 FT DOMAIN 626 659
 FT DOMAIN 661 689
 FT DOMAIN 691 727
 FT DOMAIN 729 765
 FT DOMAIN 767 803
 FT DOMAIN 806 842
 FT DOMAIN 844 880
 FT DOMAIN 882 928
 FT DOMAIN 930 966
 FT DOMAIN 968 1004
 FT DOMAIN 1006 1044
 FT DOMAIN 1046 1085
 FT DOMAIN 1087 1126
 FT DOMAIN 1130 1171
 FT DOMAIN 1173 1212
 FT REPEAT 1213 1246
 FT REPEAT 1247 1286
 FT REPEAT 1287 1326
 FT REPEAT 1327 1365
 FT REPEAT 1366 1405
 FT REPEAT 1406 1445
 FT REPEAT 1446 1485
 FT REPEAT 1486 1525
 FT REPEAT 1526 1565
 FT REPEAT 1566 1605
 FT REPEAT 1606 1645
 FT REPEAT 1646 1685
 FT REPEAT 1686 1725
 FT REPEAT 1726 1765
 FT REPEAT 1766 1805
 FT DISULFID 28 41
 FT DISULFID 35 51
 FT DISULFID 53 62
 FT DISULFID 68 80
 FT CHAIN 1467 2003
 FT CHAIN 1432 2003
 FT CHAIN 1467 2003
 FT DOMAIN 24 1447
 FT TRANSMEM 1448 1468
 FT DOMAIN 1469 2003
 FT DOMAIN 24 63
 FT DOMAIN 64 115
 FT DOMAIN 118 155
 FT DOMAIN 156 192
 FT DOMAIN 194 232
 FT DOMAIN 234 274
 FT DOMAIN 276 312
 FT DOMAIN 314 353
 FT DOMAIN 355 391
 FT DOMAIN 392 430
 FT DOMAIN 432 473
 FT DOMAIN 475 511
 FT DOMAIN 513 549
 FT DOMAIN 551 587
 FT DOMAIN 589 625
 FT DOMAIN 626 659
 FT DOMAIN 661 689
 FT DOMAIN 691 727
 FT DOMAIN 729 765
 FT DOMAIN 767 803
 FT DOMAIN 806 842
 FT DOMAIN 844 880
 FT DOMAIN 882 928
 FT DOMAIN 930 966
 FT DOMAIN 968 1004
 FT DOMAIN 1006 1044
 FT DOMAIN 1046 1085
 FT DOMAIN 1087 1126
 FT DOMAIN 1130 1171
 FT DOMAIN 1173 1212
 FT REPEAT 1213 1246
 FT REPEAT 1247 1286
 FT REPEAT 1287 1326
 FT REPEAT 1327 1365
 FT REPEAT 1366 1405
 FT REPEAT 1406 1445
 FT REPEAT 1446 1485
 FT REPEAT 1486 1525
 FT REPEAT 1526 1565
 FT REPEAT 1566 1605
 FT REPEAT 1606 1645
 FT REPEAT 1646 1685
 FT REPEAT 1686 1725
 FT REPEAT 1726 1765
 FT REPEAT 1766 1805
 FT DISULFID 28 41
 FT DISULFID 35 51
 FT DISULFID 53 62
 FT DISULFID 68 80
 FT CHAIN 1467 2003
 FT CHAIN 1432 2003
 FT CHAIN 1467 2003
 FT DOMAIN 24 1447
 FT TRANSMEM 1448 1468
 FT DOMAIN 1469 2003


```

FT DISULFID 74 103 BY SIMILARITY.
FT DISULFID 105 114 BY SIMILARITY.
FT DISULFID 122 133 BY SIMILARITY.
FT DISULFID 127 143 BY SIMILARITY.
FT DISULFID 145 154 BY SIMILARITY.
FT DISULFID 160 171 BY SIMILARITY.
FT DISULFID 165 180 BY SIMILARITY.
FT DISULFID 182 191 BY SIMILARITY.
FT DISULFID 198 211 BY SIMILARITY.
FT DISULFID 205 220 BY SIMILARITY.
FT DISULFID 222 231 BY SIMILARITY.
FT DISULFID 238 249 BY SIMILARITY.
FT DISULFID 243 262 BY SIMILARITY.
FT DISULFID 264 273 BY SIMILARITY.
FT DISULFID 280 291 BY SIMILARITY.
FT DISULFID 285 300 BY SIMILARITY.
FT DISULFID 302 311 BY SIMILARITY.
FT DISULFID 318 332 BY SIMILARITY.
FT DISULFID 326 341 BY SIMILARITY.
FT DISULFID 343 352 BY SIMILARITY.

```

Query Match 14.2%; Score 959.5; DR 1; Length 2003;

Best Local Similarity 26.5%; Pred. No. 7.1e-49; Indels 437; Gaps 64;

Matches 298; Conservative 64; Mismatches 324; Indels 437; Gaps 64;

```

QY 94 CCGFYEGSEMCVPHCADKC-----VHGRCIAPNT-----CCEPFGMGTCNSACDGDH 143
DB 105 CLRGF--TGERCAKLEDPERPSCSKRGCHQASGRPCSCMPGTGEGCQUR----- 157
QY 144 WPHCTSRCCCKNGALCNPTG--ACHCAAGFRGWCE--DRCEOG---TYGNDCHQ- 193
DB 158 --DFCSAN--PCVNGVCLATYTPQIOCHCPPEFEGHACERDVNECFODPPCPKGTSCANT 214
QY 194 -----RCQ-----CONGATC-----DHTGSCRRPPPTGAFCE 222
DB 215 LGSFQCLPYGEGPRCELRAGPCPPRGCSNGGCOLMPEKDTFHLCLCPPEGTGDC 274
QY 223 -----DLCPPGKHG-----POEORCP--CONGVCHH 248
DB 275 VNPNCVSHCCQNGTCQDGLDITTCLEPFTWGMDCSEDEVECTGPPHCRNGGTCQN 334
QY 249 VTG--ECCSPSGMWGTVCGGP-----CPEGRFGKNCSE- 280
DB 335 SAGSFHCVCYSGWGTGSCENLDICAIATCAPGSTCIDRVGSFCLCPPTGTGLCHLED 394
QY 281 -C---QCHNGTC--DAATGQ--CHCSPGTGERCQ---DECPRTGTGYLCAETGCVNG 329
DB 395 MCLSPCHGDAOCSTNLTSTLCQPGXSGPTCHQDDECLMAOQG-----PSPCEHG 449
QY 330 GKCYHVSQA--CLCEAGFAGECEAR-----LCPEGLYGIK 363
DB 450 GSCLTNTGSCFNCLCPRYTSCREADHNECLSPCHGSGTCLDLATFHCLCPGLEBGL 509
QY 364 CD---KRC---PCHLENTSHCHPMG--ECACKRGMGSLYCN----- 398
DB 510 CEVETNECASAPC--LHNAHCHDLINGFQICLPGFSGTCEBEDIDECSSPCANGGCGQ 567
QY 399 -----TCSPTGYGACQ--QIC---SCQAGACDSTYTK--CTCAGGFKIDCSTP 442
DB 568 DQPGAFHCKCLPGEFGRCTEVEDECLSDPCPVGASCLDLPGAFFCLCPGSGTGLQLEVP 627
QY 443 -----CPLSTYGI-----NCSRCG--CKNDVACSPYDGSCTC 473
DB 628 LCAPNLCQPKQICQKDKANDKANCCLPDCSPGAPEDNCTCHHGHCQ-----SSCYC 679
QY 474 KAGHGVDC-----SIRCPSGTWGFGCN---LTCQ---CL 502
DB 680 DVGTTGPECEAELEGCCISAPCAGGTCTYPPPSGYNCTCPGTGPTSCSEMTACHSPCL 739
QY 503 NGGACNTLDG--TCTCAPGRGENCE--LPC-----QDGTGLINCA----- 539
DB 740 NGGSCNPSPGGYCTCPPSHTGQCCSTYDYCVASAPCFNGGTCVNRGTPSCCLAMGFGG 799

```

```

QY 540 ERCD-----CSHADGCH--PTTGCHRCCLPGMSGVCHDS----- 570
DB 800 PROEGKLRPCSCADSPCNRRATFCODSPGPRCLCTGTGTGSSCQTDLCAQKPCPNNSHC 859
QY 571 -----VCABGRMGPNCSLP-----CYCKNGASCSPDG---ICEC 602
DB 860 LQTPSFHCLCIGWGTGPLCNLPLSSQKAALSGQIDVSSLCHNGGLC--VDSGPFYFCHC 918
QY 603 APFRTGTQCR-----ICSGFGFGHRSQRCPCQVHSSGP 637
DB 919 PPGFQSLCDHVNHPESRRCQNGATCMAPSGYLQCAAGYGQNSKRLDAC--QSP 976
QY 638 CHHTITGCDLDPFTGALCNEVCPSGRFGKNCAG-----ICTTNNGTCPNIDRS-- 687
DB 977 CHN--HGCTPRKPG--GRHC--ACRPGFVGLRCGDDVECLDQCHPTGTAAACHSANAAY 1031
QY 688 CQCYPGWTGSDCSQPCPRAHMGPNCHTCNCHNGAFCSAYDG---CKCTPGWTGLYC 742
DB 1032 CQCLPFGHTGQWCEVEIDPCHSOP-----CFHGGTCATAGSPLGFIHCPRGFEGPTC 1084
QY 743 TORCP--LGFYGNKCALICQCONGADC-----DHSQCTCRTPFMGRHC--EQKPSGTYG 795
DB 1085 SHRAPSCGFH-----HCHHGGLCLPSRPGPPPCACLSYGIGGDCULPPAPK---- 1132
QY 796 YGCRQICDCLNNSTCDHITG-----TCYSPGKRGARCDQAG 832
DB 1133 -GGPSPCLYNGSCSETTGLGGRGFRGSCPHSSPGRCKPG 1174

```

```

RESULT 14
FBPL_STRPU STANDARD; PRT; 1064 AA.
ID FBPL_STRPU
AC P10079;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Fibropellin I precursor (Epidermal growth factor-related protein 1)
DE (UEGF-1).
GN BGF1.
OS Strongylocentrotus purpuratus (Purple sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinzoa;
OC Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;
OC Strongylocentrotus.
OX NCBI_Taxid=7668;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90112459; PubMed=2514273;
RA Delgadillo-Reynoso M.G., Rollo D.R., Hursh D.A., Rafi R.A.;
RT "Structural analysis of the uegf gene in the sea urchin
RT strongylocentrotus purpuratus reveals more similarity to vertebrate
RT than to invertebrate genes with EGF-like repeats."
RL J. Mol. Evol. 29:314-327(1989).
RN [2]
RP SEQUENCE OF 279-476 AND 781-1064 FROM N.A.
RX MEDLINE=87319677; PubMed=3498216;
RA Hursh D.A., Andrews M.E., Rafi R.A.;
RT "A sea urchin gene encodes a polypeptide homologous to epidermal
RT growth factor."
RL Science 237:1487-1490(1987).
RN [3]
RP AVIDIN-LIKE DOMAIN.
RX MEDLINE=89196806; PubMed=2784773;
RA Hunt L.T., Barker W.C.;
RT "Avidin-like domain in an epidermal growth factor homolog from a sea
RT urchin."
RL FASEB J. 3:1760-1764(1989).
RN [4]
RP CHARACTERIZATION.
RX MEDLINE=91285254; PubMed=2060714;
RA Bisgrove B.W., Andrews M.E., Rafi R.A.;
RT "Fibropellins, products of an EGF repeat-containing gene, form a
RT unique extracellular matrix structure that surrounds the sea urchin
RT embryo."

```

RL Dev. Biol. 146:89-99(1991).
 CC -1- FUNCTION: FORM THE APICAL LAMINA, A COMPONENT OF THE EXTRACELLULAR
 CC MATRIX.
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR, IN VESICLES IN THE CYTOPLASM
 CC OF UNFERTILIZED EGGS, THEN TO THE BASE OF THE HYALIN LAYER
 CC THROUGHOUT DEVELOPMENT AND FINALLY IN THE APICAL LAMINA IN LATE
 CC EMBRYOS AND EARLY LARVAE.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1A (SHOWN HERE) AND 1B: ARE
 CC PRODUCED BY ALTERNATIVE SPLICING. THE SMALL FORM (1B) LACKS 8 EGF
 CC REPEATS.
 CC -1- DEVELOPMENTAL STAGE: MODERATE LEVELS IN UNFERTILIZED EGGS AND
 CC DURING EARLY CLEAVAGE, THEN RAPIDLY INCREASES IN ABUNDANCE BETWEEN
 CC LATE MORULA AND MESENCHYME BLASTULA STAGES TO MAXIMAL LEVELS
 CC MAINTAINED THROUGH SUBSEQUENT STAGES. EXPRESSED BOTH MATERNALLY
 CC AND ZYGOTICALLY.
 CC -1- SIMILARITY: CONTAINS 21 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: THE C-TERMINAL DOMAIN OF THIS PROTEIN IS SIMILAR
 CC TO AVIDIN/STREPTAVIDIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; L08692; AAA62164.1; -
 DR EMBL; L08692; AAA62163.1; -
 DR EMBL; X17530; CA33571.1; -
 DR EMBL; M17421; AAA30050.1; -
 DR EMBL; X17533; CA35573.1; -
 DR PIR; A29316; A29316.
 DR HSSP; P01132; 1EGF.
 DR InterPro; IPR000152; Axx_hydroxyl.
 DR InterPro; IPR000088; Avidin.
 DR InterPro; IPR000859; CUB_domain.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR000742; EGF_2.
 DR InterPro; IPR001881; EGF_CA.
 DR InterPro; IPR001438; EGF_IT.
 DR InterPro; IPR002049; Laminin_EGF.
 DR Pfam; PF00008; EGF; 21.
 DR Pfam; PF0431; CUB; 1.
 DR Pfam; PF01382; Avidin; 1.
 DR PRINTS; PR00010; EGFBL00D.
 DR PRINTS; PR00011; EGFBLAMININ.
 DR SMART; SM00042; CUB; 1.
 DR SMART; SM00179; EGF_CA; 18.
 DR PROSITE; PS00001; EGF_like; 3.
 DR PROSITE; PS00010; ASX_HYDROXYL; 19.
 DR PROSITE; PS00022; EGF_1; 19.
 DR PROSITE; PS00577; AVIDIN; 1.
 DR PROSITE; PS01180; CUB; 1.
 DR PROSITE; PS01186; EGF_2; 19.
 DR PROSITE; PS01187; EGF_CA; 18.
 KW Biotin; Alternative splicing; EGF-like domain; Repeat; Signal;
 KW Glycoprotein; Calcium-binding.
 FT SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 1064 FIBROPELLIN I.
 FT DOMAIN 20 55 EGF-LIKE 1.
 FT DOMAIN 62 175 CUB.
 FT DOMAIN 176 212 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 214 250 EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 252 288 EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 290 326 EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 328 364 EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 366 402 EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 404 440 EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 442 478 EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 480 516 EGF-LIKE 10, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 518 554 EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).

FT DOMAIN 556 592 EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 594 630 EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 632 668 EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 670 706 EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 708 744 EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 746 782 EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 784 820 EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 822 858 EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 860 896 EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 898 934 EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 936 1064 AVIDIN-LIKE.
 FT DISULFID 23 34 BY SIMILARITY.
 FT DISULFID 28 43 BY SIMILARITY.
 FT DISULFID 45 54 BY SIMILARITY.
 FT DISULFID 180 191 BY SIMILARITY.
 FT DISULFID 185 200 BY SIMILARITY.
 FT DISULFID 202 211 BY SIMILARITY.
 FT DISULFID 218 229 BY SIMILARITY.
 FT DISULFID 223 238 BY SIMILARITY.
 FT DISULFID 240 249 BY SIMILARITY.
 FT DISULFID 256 267 BY SIMILARITY.
 FT DISULFID 261 276 BY SIMILARITY.
 FT DISULFID 278 287 BY SIMILARITY.
 FT DISULFID 294 305 BY SIMILARITY.
 FT DISULFID 299 314 BY SIMILARITY.
 FT DISULFID 316 325 BY SIMILARITY.
 FT DISULFID 332 343 BY SIMILARITY.
 FT DISULFID 337 352 BY SIMILARITY.
 FT DISULFID 354 363 BY SIMILARITY.
 FT DISULFID 370 381 BY SIMILARITY.
 FT DISULFID 375 390 BY SIMILARITY.
 FT DISULFID 392 401 BY SIMILARITY.
 FT DISULFID 408 419 BY SIMILARITY.
 FT DISULFID 413 428 BY SIMILARITY.
 FT DISULFID 430 439 BY SIMILARITY.
 FT DISULFID 446 457 BY SIMILARITY.
 FT DISULFID 451 466 BY SIMILARITY.
 FT DISULFID 468 477 BY SIMILARITY.
 FT DISULFID 484 495 BY SIMILARITY.
 FT DISULFID 489 504 BY SIMILARITY.
 FT DISULFID 506 515 BY SIMILARITY.
 FT DISULFID 522 533 BY SIMILARITY.
 FT DISULFID 527 542 BY SIMILARITY.
 FT DISULFID 544 553 BY SIMILARITY.
 FT DISULFID 560 571 BY SIMILARITY.
 FT DISULFID 565 580 BY SIMILARITY.
 FT DISULFID 582 591 BY SIMILARITY.
 FT DISULFID 598 609 BY SIMILARITY.
 FT DISULFID 603 618 BY SIMILARITY.
 FT DISULFID 620 629 BY SIMILARITY.
 FT DISULFID 636 647 BY SIMILARITY.
 FT DISULFID 641 656 BY SIMILARITY.
 FT DISULFID 658 667 BY SIMILARITY.
 FT DISULFID 674 685 BY SIMILARITY.
 FT DISULFID 679 694 BY SIMILARITY.
 FT DISULFID 696 705 BY SIMILARITY.
 FT DISULFID 712 723 BY SIMILARITY.
 FT DISULFID 717 732 BY SIMILARITY.
 FT DISULFID 734 743 BY SIMILARITY.
 FT DISULFID 750 761 BY SIMILARITY.
 FT DISULFID 755 770 BY SIMILARITY.
 FT DISULFID 772 781 BY SIMILARITY.
 FT DISULFID 788 799 BY SIMILARITY.
 FT DISULFID 793 808 BY SIMILARITY.
 FT DISULFID 810 819 BY SIMILARITY.
 FT DISULFID 826 837 BY SIMILARITY.
 FT DISULFID 831 846 BY SIMILARITY.
 FT DISULFID 848 857 BY SIMILARITY.
 FT DISULFID 864 875 BY SIMILARITY.
 FT DISULFID 869 884 BY SIMILARITY.
 FT DISULFID 886 895 BY SIMILARITY.
 FT DISULFID 902 913 BY SIMILARITY.
 FT DISULFID 907 922 BY SIMILARITY.

FT DISULFD 924 933 BY SIMILARITY.
 FT CARBOHYD 30 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 136 136 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 851 851 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARSPLIC 477 780 MISSING (IN ISOFORM IB).
 FT CONFLICT 279 279 L -> S (IN REF. 2).
 SQ SEQUENCE 1064 AA; 112072 MW; 2E569CA012ED6D09 CRC64;

Query Match 14.2%; Score 954.5; DB 1; Length 1064;
 Best Local Similarity 28.0%; Pred. No. 7.6e-49;
 Matches 290; Conservative 93; Mismatches 305; Indels 347; Gaps 69;

QY 30 DPNVCSHWSYVTVQESYRPHFDQIYTTCTDILNMFKCTHRVSYRTAYRHGKMTMR 89
 DB 191 DPNLQNG-----ACCTDLYNDACT----- 201
 QY 90 RKSQCCPEFYSEGENC---VPHCA-DKCVHGRCLAPN---TCQEPGWTGNCSSACDG 141
 DB 202 ---CPGPF--TGRNCEIDIDEACADPCONGACVGVNGVYCNCPGFDGECENIN- 254
 QY 142 DHMHPCHSRCKNGALCNPTGA-----CHCAAGFRGMRCE---DREGGTGNDCHOR 194
 DB 255 -----ECAS-SPCLNGGIC--VDGVNMECTCTAGFTGRCVFNIDECAS----- 296
 QY 195 GCGCQNGATC-DHYTG-ECRCPGYTGAFCEDLCPGKHGPOEORCPONGGVCHHYTGE 252
 DB 297 APCONGGICIDINGINTYCTSCPLGFSGDNCE-----NDDECS--IFCLNGTCTVDLNA 349
 QY 233 ---CSCPSSMGTVCQPCPEGRFGKNGSQECCNGHTC-DAATG-QCHSPGYNGERCO 308
 DB 350 YMCVAPMTGFTCADNIDE-----CA-SAPCQNGVCIDGVNMYMDCQPGYGTGHE 402
 QY 309 ---DECPVGTGVLCAETCCQVNGGKYH-VSG-ACICEAGRAGRCARLCBELYGLK 363
 DB 403 TDIDEC-----APPCONGGDCVGVNGVYCTCAPGFDGNCE----- 440
 QY 354 CDKRCPCHLENTHSC--HPMSGECACKPGWSGLYCNETCSPGYEACQ---QICS---C 415
 DB 441 -----NNIDECASRPCQNGAVCDVNGVFC--TCSAGYTVLCTDIDECASMPIC 489
 QY 416 ONGADC-DSYTGK-CTCAPGFRGIDCSTPCLGTYGINSFRGCKKNDAYCS-PVDG-SC 471
 DB 490 LMGVCTDLVNGYICTCAAGEGTGCTD-----ECAS-FPCQNGATCTDQVNGYVC 542
 QY 472 TCKAGMWGDC-----SIRCPSG-----TWGFGCNL-----TCQ----- 500
 DB 543 TCVPEYTVLCTDIDECASFPCLNGGTCNDQVNGVYCAADIVSTCTEDRDECASAP 602
 QY 501 CLNGGAC-NTLDG-TCGAPGWRGKCEI-----PCQDGTGILNCAERDCSHADGC 550
 DB 603 CLNGGACADVNGVFCYCTCLPMEGTNCEINIDECASSPCQNG--GL-CYDQVN-SYV--- 655
 QY 551 HPTTHCRCLPGWSGVHCDSCVACBRWGPNCSLPYCKNGASC--SPDGCICECAPFRG 608
 DB 656 -----CCLPGFTGIHGTEIDECASSP-----CLNGGQCIDRVSYECVCAAGYTA 702
 QY 609 TTQO-----RSCSPGYHRSQGTQPCQVHSGPC----- 638
 DB 703 VRCQINIDECASAPONGCVGVNGVYCNAPGTYGDNCTEIDE--ASMPCLNGCA 760
 QY 639 ---HHITG-LCDCLPGFTGALC-----NEVCPGSRFGKNGAGICTCTNNGTCPIDRS 687
 DB 761 CLEWNGYTCQCVAGYTVICTDIDECASAPCQNG-----GVCITDININ- 807
 QY 668 CQCTGWMIGSDCQPCPAHMGPNCTHPCNCHNGAFCSAYDG---ECKCPGWTGLYCT 743
 DB 808 CACVGFSGNSCETNIDECASDP-----CLNGGIC--VGVNGVFCQCPNYSGYTCE 858
 QY 744 QRCPLGFGKCALICQONGADCDHISQ--CTCRTGPMGNHC---QKCSGYTYGVC 798
 DB 859 -----ISLDACRMPQNGATCVNAGADYCECVPGAGNCEIDIDECAS----- 904
 QY 799 ROICDCLNNSTC-DHITG-TCYCSPGMKGARCDAQGIIVGNL-----NSLSRTSTA 848

DB 905 ---LPCQNGGICIDIDAGYTCQCRGLGYGVNCEEVGFCDLEGMWYNECNDQYTIKSTG 961
 QY 849 LPADSYQIALAGIITIVLVVLFLLALFTIYHKKRGKRESSPAPATYTPARVYNAADTTI 908
 DB 962 M-----MLGDWYTNERALGYAAPVYVYASN-----NYDPPS 995
 QY 909 SG-TLPHSNGGNANS 922
 DB 996 FGVTVVRDNGOSTTS 1010

RESULT 15
 NTCA_MOUSE STANDARD: PRT: 1964 AA
 AC P31695; Q62389; Q62390; Q35442; Q9RLW9; Q88314; Q88316; Q9RLX0;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Neurogenic locus notch homolog protein 4 precursor (Notch 4).
 GN [contains: Transforming protein Int-3].
 OS NOTCH4 OR INT3 OR INT-3.
 ON Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92194507; PubMed=1312643;
 RA Robbins J., Blondel B.J., Gallahan D., Callahan R.;
 RT "Mouse mammary tumor gene Int-3: a member of the notch gene family
 transforms mammary epithelial cells."; J. Virol. 66:2594-2599(1992).
 RL [2]
 RN REVISIONS, SEQUENCE FROM N.A.
 RX MEDLINE=97294599; PubMed=9150355;
 RA Gallahan D., Callahan R.;
 RT "The mouse mammary tumor associated gene INT3 is a unique member of
 the NOTCH gene family (NOTCH4)."; Oncogene 14:1883-1890(1997).
 RL [3]
 RN SEQUENCE FROM N.A.
 RP TISSUE-Lung, and Testis;
 RX MEDLINE=96281668; PubMed=8681805;
 RA Uytendaele H., Marazzi G., Wu G., Yan Q., Sassoon D., Kitajewski J.;
 RT "Notch4/Int-3, a mammary proto-oncogene, is an endothelial
 cell-specific mammalian Notch gene."; Development 122:2251-2259(1996).
 RL [4]
 RN SEQUENCE FROM N.A.
 RP Roken L., Mahairas G., Qin S., Ahern M.E., Dankers C., Lasky S.,
 RA Loretz C., Schmidt S., Tipton S., Traicoff R., Zackrone K., Hood L.;
 RT "Sequence of the mouse major histocompatibility locus class III
 region."; Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
 RL [5]
 RN SEQUENCE OF 1436-1600 FROM N.A.
 RX MEDLINE=99252212; PubMed=10233982;
 RA Lee J.-S., Haruna T., Ishimoto A., Honjo T., Yanagawa S.-I.;
 RT "Intracisternal type A particle-mediated activation of the Notch4/Int3
 gene in a mouse mammary tumor: generation of truncated Notch4/Int3
 mRNAs by retroviral splicing events."; J. Virol. 73:5166-5171(1999).
 RL [6]
 RN FUNCTION.
 RX MEDLINE=21244657; PubMed=11344305;
 RA Uytendaele H., Ho J., Rossant J., Kitajewski J.;
 RT "Vascular patterning defects associated with expression of activated
 Notch4 in embryonic endothelium."; Proc. Natl. Acad. Sci. U.S.A. 98:5648-5648(2001).
 RL [7]
 RN SEQUENCE OF 1463-1964, POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS
 OF VAL-1463.


```

OY 95 CP-GEYEGMCPVPCADKCY-----HGRCIAPT-----COCBPGMGNGNSACDGDH 143
Db 102 CPBGF--TGDKQTHLELCLCPSPFCNSMGHCYVQASPPCOSCPEWGTGBOCLR-----154
OY 144 WGPHTSRCKCKNGALCNPTIG--ACHQAGFRGMCE-----DCEOGTYGNDG 191
Db 155 --DPCSAN-PCANGVCLATYPOIQCRCPREFBHTCERINCEFTLEPGCCPGT---SC 208
OY 192 HQ-----RCQ-----CONGATCD-----HYTGE-CRCPGYTGA 219
Db 209 HNTLGSYQCLCPVQEGPQCKLRGACRPBGLNGGTQCLVPEBHSFTHLCLCPGPTGL 268
OY 220 FCE-----DLCPRGKHG-----POCEORCP--CONGV 245
Db 269 DCEMNPDDCVRHOCONGATCLDGLDITYTCLCPKTMKMGDCSEDIDECARGPPRCRNGT 328
OY 246 CHVYTG--ECSCPSGMWKTVGCP-----CPEGRFGKNS 278
Db 329 CONTAGSFHCYCVSGMWGAGCEBNDCAATCARSTCIDRVGSFSLCPRGRTGLCH 388
OY 279 QE--C--QCHNGSTC--DAATGQ--CHCSPGYGERCO--DECPVGYTVLCAETQOC 326
Db 389 LEDMCLSQPCVHNAQCSNPLTGSTLICQPGYSGSTCHODLDECOQAQOG-----PSPC 443
OY 327 VNGGCKYVSGA--CLEAGFAGERCEAR-----LCPEGLY 360
Db 444 EHGGSCLNTPGSEFNCLCLPGYTGSRCEADHNECLSDPCBPGSTCLDLATFHCLCPGLE 503
OY 361 GIKCD--KRC--PCHLEHTSCHPMMSG--ECACKPGMSGLYCN-----398
Db 504 GRLEVEVNECTSNPC--LNOACHDLNGLFQCLCLPFTGARCEKMDDECSTPCANGG 561
OY 399 -----TCSPEFYEGACQOIGS-----CONGADCDVYTG--CTCAPGFKGIDC 439
Db 562 RCRDOPGAFYCECLPFGFGPHCEKEVDECLSDPCPVAGASCLDLPGAFFCLCRPGFTGLC 621
OY 440 STP-----CPLGTGT--NCSBRCG-----457
Db 622 EVPLCTPNNMCPGQOCQOGEHRAPLCLPDSPGCVPAEDNCPCHHGHCQRLCYCDEGWT 681
OY 458 -----CKNDVAVCSVPDG--SOTCKAGMHGVDCS--IRCPSGTW-----491
Db 682 GRECTETELAGGCTSPCAHAGTCHPQPSGYNCTCPAGTMGLTCEBETVACHSGPCLNGSGC 741
OY 492 -----GFGCN-----LTCQCLNGGACNTLDGT--CTCAPGRGEXCE- 526
Db 742 SIRPBGYSTCLPSTHGRHCOTAVDHCVASCLNGGTCVKNPGETFFCLCATGFGHLCEB 801
OY 527 -----LPCODGTYLNC-----AERCDCSHADCHPT-- 553
Db 802 KTNPSCADSPRNKATQDTPRGARCLCSPGYTSSCQTLIDLARPCPHTRARCLQSGP 861
OY 554 TGHGCLGPGMSGVND--SYCABGRMGNCSLPYCKNGASCSPPDG--ICECAPBFRG 608
Db 862 SFQCLCLGWTGALCDPPLSCQAMASOGIEISGLCONGGCLT-DTGSSTYFCRCPPFGOG 920
OY 609 TTCORICSPGFYGRHCQTCPCQVHSSGPCNHTITGLCDCLPGFTGALCNEVCSGRFGKN 668
Db 921 KLCODNNMP-----C--EPNCHHS--TCVPQPSGYVCO--CAPGYEGON 960
OY 669 CAGT--C--TCTNNGTG--NPIDRSQCYPGWIGSDC-----SDPCPAHMGPMC 712
Db 961 CSKYLEACQSPCHNHGCTSRPGGFHCACRPFGVGLRCEGDVDECLDRCHPS-----1014
OY 713 IHTCHNGAGFCAVDECKCTPGWTGLYC-----TQ 744
Db 1015 -GTAAH--SIANAF--YCQCLPGHTGQRCVEEMDLQSQPCSNNGSCETITTPPFGTFC 1069
OY 745 RCPLEFYKDC--ALIC--QCONGADC-----DHISGQCTCTGFMGRHC--EQKCPSG 792
Db 1070 HCPKGFEBPTSHKALSGIHHCHNGGLCLPSRPGSPPLACLSGFGGPDCLTPPAPP- 1128
OY 793 TYVGCRQICDCLNNSTCDHITG-----TCYCSPGMKGARCDOAG 832

```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 9, 2003, 11:47:55 ; Search time 33.0243 Seconds

(without alignments)
3318.560 Million cell updates/sec

Title: US-10-092-390-2

Perfect score: 6744

Sequence: 1 MVLINSLSCSLFILLCHMT.....SSPKQDSSGSSSSSSSSSS 1140

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 segs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

1: PIR-73:*
2: PIR1:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1958	29.0	1574	2 T13954	MEGF6 protein - ra
2	1900	28.2	1620	2 T27283	hypothetical prote
3	1805.5	26.8	1111	2 T26972	hypothetical prote
4	1036	15.4	2524	2 A35844	Notch protein - Af
5	1032	15.3	2555	2 A40043	Notch protein homo
6	1028	15.2	2531	2 A46019	Notch-1 protein -
7	1024	15.2	2531	2 A18188	Notch protein homo
8	998	14.8	2471	2 A49128	cell-fate determin
9	987	14.6	2437	2 S42612	transmembrane prot
10	978.5	14.5	2703	1 A24420	Notch protein - fr
11	977.5	14.5	2318	2 S45306	Notch 3 protein -
12	974	14.4	2321	2 S78549	Notch3 protein - h
13	964.5	14.3	2531	2 T31070	Notch homolog - se
14	954.5	14.2	1064	2 A40136	fibropellin Ia - s
15	952.5	14.1	1964	2 T09059	Notch B protein -
16	939.5	13.9	1203	2 A49175	Notch B protein -
17	920.5	13.6	4006	2 T09070	probable tenascin
18	915.5	13.6	2352	2 T30201	Notch homolog prot
19	914.5	13.6	3566	1 A40701	tenascin-X precurs
20	881.5	13.1	4135	2 T42629	tenascin-X - bovln
21	775.5	11.5	2139	2 A35672	tenascin-C - human
22	769	11.4	2201	2 A32160	tenascin-C - human
23	767.5	11.4	1746	1 S19694	tenascin precursor
24	746.5	11.1	2019	1 J01322	tenascin precursor
25	744	11.0	1220	2 A56136	jagged protein pre
26	720.5	10.7	861	2 A48825	Notch homolog Motc
27	717.5	10.6	1801	1 MMR5	laminin beta-2 cha
28	716	10.6	1801	1 S53868	laminin alpha-2 ch
29	709.5	10.5	1810	1 A32230	tenascin precursor

30	704.5	10.4	3084	1 MMSA	laminin alpha-1 ch
31	695.5	10.3	1790	1 MMR1	laminin beta-1 cha
32	692	10.3	3075	2 S14458	laminin alpha-1 ch
33	690	10.2	1408	2 S16148	gene serate prote
34	685.5	10.2	1798	2 S53869	laminin beta-2 cha
35	683.5	10.1	3635	2 T10053	laminin alpha 5 ch
36	676.5	10.0	3672	2 T32433	hypothetical prote
37	676.5	10.0	3704	2 T37316	probable laminin a
38	671	9.9	1722	2 E89753	protein F1C7.4 (1
39	667	9.9	1766	1 MMH01	laminin beta-1 cha
40	666	9.9	2823	2 T23064	hypothetical prote
41	666	9.9	2823	2 F87908	protein T22A3.8 (1
42	666	9.9	3102	2 T43291	laminin alpha cha1
43	664	9.8	1786	1 MMSB1	laminin beta-1 cha
44	657	9.7	1429	2 S06434	homeotic protein 1
45	641	9.5	1797	2 A55677	laminin beta-2 cha

ALIGNMENTS

RESULT 1

T13954

MEGF6 protein - rat

C:Species: Rattus norvegicus (Norway rat)

C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 21-Jul-2000

C:Accession: T13954

R:Nakayama, M.; Nakajima, D.; Nagase, T.; Nomura, N.; Seki, N.; Ohara, O.

Genomics 51, 27-34, 1998

A:Title: Identification of high-molecular-weight proteins with multiple EGF-like mot.

A:Reference number: 214126; PMID:98360089; PMID:9693030

A:Accession: T13954

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1574 <MAX>

A:Cross-references: EMBL:AB011532; NID:93449293; PIDN:BA032462.1; PID:93449294

A:Experimental source: strain Sprague-Dawley; brain

C:Genetics:

A:Gene: MEGF6

Query Match	29.0%	Score 1958	DB 2	Length 1574
Best Local Similarity	41.3%	Pred. No. 1.1e-96		
Matches 344	Conservative 77	Mismatches 306	Indels 106	Gaps 16
QY	95	CP-GFYESGECVP--HCADKCVHGR--IANTQCEFGMGTCSSACDSDHMGPHCTS	150	
DB	602	CPKGFY--GKRCRRKCHCANR--GRCHRLYGACIDPGLGRCHLACPPMAFGPGSE	656	
QY	151	RCOCKNG--ALCNPTTGACHCAAGFRGWRCEDEOGYGNDCROROCOGATCDHYTG	208	
DB	657	DCLEQSHNSCNPKDSCSCSKAGFQGBRCABECSGFFGCRHRTCDPGVACDPVSG	716	
QY	209	ECR--CPFGYTGAFCEDLCPPGKHGPOCEORPCONGVCHHVTGEGSCSPSGMGTVCQ	266	
DB	717	ECRQCCPPYGGEDCGQCEPVGTGVNCSGSCSV--GAPCHRYAGEICLCPKGTGEBOGA	775	
QY	267	PCPEGRGKNCQRC--OCHNGGTDAATGQCHSPGTGEGCEQCEPCRYGYVLAETCQ	325	
DB	776	DCPEGRWGLGQCEICPACCHGASCNPETGTCLDPFGVGRSCDPTCSAGWGTCCQIRCA	835	
QY	326	CVNCG--KCYH-----VSGAICG	341	
DB	836	CANDHCDPTTGRCSCAPGWTGLSCQACDSGHMGPPDIPHCNCSAGHNCDAVSGLCIC	895	
QY	342	EAGFAGECEARLCPESLYGICDKRCPCPLHENTHSCHPMSGECACRPGSGLYCNETCS	401	
DB	896	EAGFAGEPRCE--QSCRGYGYGSPCEQKCR--EHGAADHVSAGACTCPAGMNGSCEHACP	952	
QY	402	PGFYG-----EA-----CQITCSQNG	418	
DB	953	AGFGGLDCDSACNCSAGAPCDAYTGSCICPAGRWGPRCAOSCPPLRTGLNCSQICCTCFNG	1012	
QY	419	ADDSVYTGKCTCAGPFGKIDGSTPCPLGTGYGINCSRCGCKNDVAVCSPVDSCTCKAGWH	478	

Db 1013 ASCDSTTGCGCHAPGMMGPTCLQACPPGLYGRKNCOHSCCRNGGRDPTILAGCTCPREGWT 1072
 QY 479 GVDSCIRCPSTGTMGFCNLTCCOLINGACNTLDGTCTCAPGRNGEKECELPCDDGTGLMC 538
 Db 1073 GLACEHECLPRGHYAMACQNLNCSLHGGIDRLTGLHCLCPAGMTGDCQSSVSGTFGVNC 1132
 QY 539 ABRCDSCSHADGCHPTTGHCRCRLPMSGVHCDVCAEGRMGNPCSLPCYKNGASCSBPDG 598
 Db 1133 EEHCAKRGKASCHHYTGACFCPPRMGRPHCEQACPRGMFGEMACAOQCLCLPTNASHHYTG 1192
 QY 599 IEBCAFGFGTTCQRCISGFGYGHRSQTCRQCVHSSGGCHHTGLCDLDPFTALCNE 658
 Db 1193 ECRCPGFGTGLSCQACQDGTGKDCENLC-OCPETMACDPAASGCTCAAGHGTGCLQ 1251
 QY 659 VCPSGFGRKNCAGICTCTNNGTNPIDRSQCQYPGWIGSDCSOPCPRAHMGPCITNC 718
 Db 1252 RCPSGYGRGCEHICLCLNGTCIDPATGACYCPAGFLGADCSLACQGRFGRSCHAIVC 1311
 QY 719 HNGAFCSAYDECKCTPGWTGLYCTQRCPLGFGYKDCALICQOQNGADCDHISGQCTRT 778
 Db 1312 RCGACDPVSGACISCPGKTGVNCEHGCQPDREFGKCELCACRNGGLCHATNGSCSCL 1371
 QY 779 GEMGRHCEKCPSGYGVYCGRCQICLCLNSTDHITGTCTYCSPGMKGARCDQ 831
 Db 1372 GMMGRPCEHACPRAGYGAACLLCEFCQNNNGSEPTTGACLGFGFYQACENS 1424

RESULT 2

727283

hypothetical protein Y64G10A.f - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T27283

R:Alnough, R.

submitted to the EMBL Data Library, September 1999

A:Reference number: 220336

A:Accession: T27283

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1620 <WIL>

A:Cross-references: EMBL:AL110498; NID:el542303; PIDN:CAB54471.1; CESP:Y64G10A.f

C:Genetics:

A:Experimental source: clone Y64G10A

A:Gene: CESP:Y64G10A.f

A:Introns: 77/1; 116/1; 198/1; 282/1; 365/1; 425/1; 466/1; 548/1; 559/1; 601/1; 625/1; 7

Query Match

Best Local Similarity 38.6%; Score 1900; DB 2; Length 1620;

Matches 331; Conservative 88; Mismatches 307; Indels 132; Gaps 17;

QY 95 CP-GEYSEEMCVPHCADXVHGRCIAP-NTQCEPFGWGTNCSACDDHMGPHCTSRG 152
 Db 712 CPDGFY--GSCNLMKRCMPCNRCDPVFGYCTCPDGLYGQSEKRCPTHFTGKNCRCPC 769
 QY 153 QC--KNGALCNPTTGACHAAGFRGRCEDRCEOGTYGDCDHORCCQNGATCDHTGEC 210
 Db 770 KCARESEGCDEITGCKRCRPGYGHCKRMCSPLGFGACMAKCSPPAGIRCDPYTGDC 829
 QY 211 --RCPGYGACFCECDLCPGKHGPOCEORPC-----QNGVCYHVTGSCSPSGMGT 262
 Db 830 TKKCRAGYGNLCLDPCRPAGYFGYDEQKCSADVASPHKSKYCHNVTGTCPLPCKTRP 889
 QY 263 VCGQ-----PCPEGRFGKNCSEOCCHNGTCDATAATGCHCSPGYT 303
 Db 890 LDDQCLIFETIEFDIASINVIACAPNTYGRPCAHCTSCVNAKAKDESDGSHCTPFRY 949
 QY 304 GERDCDECPVGYVCACTCOCVNGKCYHVSAGCLCAGRAGERCEARLCPREGLYGK 363
 Db 950 GATSCVCTPGRFGIDCKLCKONGAICDTSNGSCCAPSGSKKD--KACAPGTGFKD 1008
 QY 364 CCKRCCHLENTHSCPHMGECAKCRKMGSLYCNETCSGFGFGEACQQLCSCQNGADCS 423

Db 1009 CSKRCDC-ADGNH-CDPSDEGECIPRCKGRHKRCDETCDSGLRGAGKGCISQNGATCDS 1066
 QY 424 YTGKTCAPGFGIDSTPCPLGTGYINCSSRGC--KND-----AVCSPYVDSCTC 473
 Db 1067 VTGSCCECRGMKCKDCRCRPGREBEGCNALCDDTTTNDTSMYNPFAVARCHVTECC 1126
 QY 474 KAGMHGVDSCIRCPSTGTMGFCNLTCCOLINGACNTLDGTCTCAPGRNGEKECELPCD-- 531
 Db 1127 PAMTGPDCQTCPLRGHGEGRHSCQCSNGASCDRVTFCCPSPGFMKNGESECPEGL 1186
 QY 532 -----CTGLNCAERDCSHADGCHPT 553
 Db 1187 WGSNCKHCLNHGGECONENGDECIDGTGTPSLCPFOFGRNCAQRNCNKGASCDRK 1246
 QY 554 TGHCRCLPMSGVHCDVCAEGRMGNPCSLPCYKNGASCSBPDGICEAPFRGTTCOR 613
 Db 1247 TGRCECLPMSGVHCDVCAEGRMGNPCSLPCYKNGASCSBPDGICEAPFRGTTCOR 1306
 QY 614 TCSPGFYGRGCEHICLCLNGTCIDPATGACYCPAGFLGADCSLACQGRFGRSCHAIVC 661
 Db 1307 PCLKGYGRHCSQSC-RCANSKS-CDHISGRQCPRGYAGHSTELCPDGTGSESCQK 1364
 QY 662 -----GGRGKNCAGICTCTNNGTNPIDRSQCQ 690
 Db 1365 DCGENSMCDALISGKCFCKPGRHSGSDKSGCVGRREBPDNQLOSCENGCVDSSTGSCVC 1424
 QY 691 YPGWIGSDCSOPCPRAHMGPCITNCNCHNGAFCSAYDECKCTPGWTGLYTORCPLGF 750
 Db 1425 PPGYISTKCEINAOQSDRFPCTCEKICNGENGGTCDRLTGQCHCLDPFTMTCONVPEBR 1484
 QY 751 YGRKDCALICQOQNGADCDHISGQCTCTRTGFMGRHCEQKCPSETGYGCAQICDCLNSTRG 810
 Db 1485 FVAGCEKRCRANG--HCNASSSECKNCLGFTGPSQBQSPGRKYGINCTLDECYQARC 1543
 QY 811 DHTGTCTYCSPGMKGARC 828
 Db 1544 DPVQGCCDPCPRGRYSRC 1561

RESULT 3

726972

hypothetical protein Y47H9C.4 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Mar-2000

C:Accession: T26972

R:Harris, B.

submitted to the EMBL Data Library, October 1998

A:Reference number: 220293

A:Accession: T26972

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1111 <WIL>

A:Cross-references: EMBL:AL032657; PIDN:CAA21739.1; GSPDB:GN00019; CESP:Y47H9C.4

A:Experimental source: clone Y47H9C

C:Genetics:

A:Gene: CESP:Y47H9C.4

A:Map position: 1

A:Introns: 50/2; 84/2; 150/1; 238/3; 342/3; 797/1; 851/1; 947/2; 1017/1; 1083/1

C:Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homol

Query Match

Best Local Similarity 26.8%; Score 1805.5; DB 2; Length 1111;

Matches 373; Conservative 162; Mismatches 423; Indels 215; Gaps 40;

QY 21 GTASPLNLEDPNVCSHWEYSVTVOESYHPPDOIYT-----SCTILMMFKTRRR 73
 Db 35 GTTEP---QGDHYCT-----VKTIVDDY--ELKAVIHYVYVYDTEGCLNPLGFGC---- 80
 QY 74 VSYRTAVRHGRTMYRRK-----SQCCPGYSEGE--MCYPHCAKDCVHGRCIAPNTC 124
 Db 81 ----TYEKRGQKASVYRQLVYKKEKYKQCCDDGYQKDFCLPCDPCPCKKCKCIPIRGK 136
 QY 125 QCEPGWGTNCSACDDGDMHGPCHTSRCQCKNGALCNPTTGACHCAAGFRGMRCEDRCHQ 184

Dd	137	EODPEGGKRYCASSSVGTMLGSKSDCEDENGMNCORDELGTCLCTSGPFGEGEKEKPCPD	196
Oy	185	GTYGMDCHQRCQCCQNGATCDHYTGECRCRPGYTGAFECDELCPGPKHNPPOEORCPQONG	244
Dd	197	NKMFGNCYKSCPCQNGGCKNK - EGKCYSDMGEEFCLNKEBKGFECKPFCNCONGA	255
Oy	245	VCHNHTGSCSPSGMMGYGQCPREGRFGKNCSEOEQCHNGGTCDAATQOCHCSGTYG	304
Dd	256	TCDDNTNGCICKSGHGLACNEBSVGFPSGGCTQKDCJLNNONCDSSECKCT - GWTG	314
Oy	305	ERCODECVGTGYVLCAETCCQV - - - - - NGKCYAHYSGALCEAGFAGERCARLCPEG	358
Dd	315	KHCDIGSGRGRHPTQCKQNCJCPGLEFSDSNASDANTGQCQCESGYKPKERCKODAE	374
Oy	359	LYGICDKRCPCHELENTHSCHPMSBECACPKFMSGLYCNETCSBFGYGEAC - - QOTCSQ	416
Dd	375	QYGADCSYTCYCBENLTMCPNTGFCRCKRGEFGDNCELACSDSYSPGPKCEKQAMDMN	434
Oy	417	NCADDSYTGKCTOAPRGKJGDCSTPRLGYGNGCSSRGCKNDAY - CSYPVGSGTCXA	475
Dd	435	HASEBNPTGSCYCKPRGTGNCSEBPPRLDYERGNCHQOCCONRGYGCSDAGDKCQDR	494
Oy	476	GMHGVDCSIRCPSGTWGFGCNLTQCLNGACNTLDGTCTCARPMREKCELPQODGYG	535
Dd	495	GWTGRCHEHNPADPTFGANCKRCKCPKJGICDPIETGECSPAGLQANCDICCPBESYG	554
Oy	536	LNCAERCCSHADGCHPTTGRCRLPMSGVHODSVABEMRNCSLPCTCNNGASCSP	555
Dd	555	PECKLHCACVNGK - CDKETGCTOPGRFGSDGCTYSKRYGSCSELSPCSD - ASCK	612
Oy	596	DDGICECAPRGRTTCORICSPGFGYGHRCQSOTQPCVHSSGPHNI - - - - - TGLC - DCLPGF	651
Dd	613	QTGKCLCPLGTGKGVSCDQKDPNTEFGLQGTV - - - - - TSPCASPDPKNGVCLSPGSG	657
Oy	652	TGALCNEVCPSBGRGKNCAGICTCTNNGTCPNIDRSQCTYPMGJGSDSPCSPRAHGMN	711
Dd	712	CIHTC - NCHNAFCSAYGDECKCPRTWGTGYCTORCEPLGFGYKCALITCCONG - ADGDH	769
Oy	728	CALDOPKASSTODHINGLCICRAGLEGALCTPCRCAAGTWGNCNRVCYKCTSEYKQCA	787
Oy	770	ISGQCTCTGTMGRHCEQKCPSGTYGIGCROICDCLNNT - - - - - CDHITGTGCSPGWKGAR	827
Dd	788	QTEGSCSPAGQGBRCDPCPEDGYGDDCIKKCKQGTAASSCKRWVSGACHHPGFTGEF	847
Oy	828	CDQAGVILVGNLSLRST - - - - - ALP - - - - - ADYQIAGALG - - - - -	861
Dd	848	C - - - - - HALCPSTIGLCKSEBPCDGGDDYECDDALIGCHNDOMSCGAKOE	896
Oy	862	- - - - - IILVLVLF - - - - - LLAETIYRHK - QKGESSMBAVYTPAMRV	901
Dd	897	FEALGAGRSTGLTWFFLLITLVALCGGILGIALF - - - - - YRNKYQKEKDPMDTYSF - - - - -	948
Oy	902	VNADVTISGTLPHNSGGMANSHYTNYSHTLQOACSPHVNRRDRTYTKSKNNOLFVN	961
Dd	949	- - - - - HAPNNDREGRQNPY - - - - - SQQVFP - - - - - OSDAFSSNNQ - - - - -	986
Oy	962	LKNVNGKRGVGDCTGTLPLDMHNGVYLELAFGLDRSYM - - - - - GKSLDLGKN -	101
Dd	987	- - - - - GPPN - - - - - GLITLEBELENKKTHGSAAGRGNDY	101
Oy	1014	- - - - - SEYSSNCSLSSSE - - - - - NPAYTICKDPVLLPKSSEGYEMK - - - - - SPA - - - - -	105
Dd	1020	ASLDEVAGESSSSASASARGLNSSEQSRP - - - - - LLEEHDEEFDEPHENSIPAHAYT	107
Oy	1056	- - - - - RDSPYAEINN - - - - - STSAR - - - - - NVY	1075
Dd	1078	TSNNHENPYADISSPDPYQNSAKKKAQDNLY	1110

Db 623 NGYICPKGTTGVNCEIKIDDCASNLCIDNGKIDKIDGIECTCPGTGKLCININIEC 682
 QY 486 -----CPSTGWGFC--- 495
 Db 683 DSNPCRNKGCTCKDOINGETCYCPDGYHDHMCISEVNECNSNPCIGHACHDGVNGYKCDCE 742
 QY 496 -----NLTCQ---CLNGACNTLDGT---CTCAPRGNGECCEL-----PC-Q 530
 Db 743 AGHSSNCNDINNNECESNPCNMGCTCKDMTGAYICTCAGFSGPCNQTINIECSSNPLCN 802
 QY 531 DGT-----YGLNC-----AERCD-----CSHADGHPY---TGHCRLPGMS 564
 Db 803 HGCTIDVAGYKCNMCLPYTGALCEAVLAPCAGSPCKNGKNGKEDEFTFSCCEBPQMO 862
 QY 565 GVHDD-----SYCABGRNGPCNSL-----PCYCANASC 593
 Db 863 GGTCEIDMNECVNRCNKGATCQNTNGSYKCNCKRPGYTGRCNCEMDIDDCQPNPCNNGSC 922
 QY 594 SPDDGI-----CECARGFGTTCQR-----ICSPGYGHC 624
 Db 923 S--DGINMFCNCRAGFRGPCEDINECASNPCKNAGCTDCVNSTYCTCQPGFSGTHC 980
 QY 625 SQTCPQCVHSS---GPCNHTGL---CDCLPGFTGALC---NE----- 658
 Db 981 ESNTPDCTESSCFNGGTC--IDGINTFCQCPGFTGSCQHDINECDKPLNGTQOD 1038
 QY 659 -----VCPSPGRGKNGAGI-----CTCTNNGTG---NPIIDSCQCTPMTISDCSQP 702
 Db 1039 SYGTCTKCTPGGYGLNCONLVNRCDSPPCKNGKRCWQTNFYR--CEKSSMTQVYCDVP 1097
 QY 703 ---CPRA--HWPNCIHTCNHNGAF--SAVDECKCTPGMTGLYCTORCPLEFYGKDC 755
 Db 1098 SVSEVAVAKOOGVDIVHL--CRNSGMCVDIGTHTFCRCQAGYTSYCEQY-----DEC 1149
 QY 756 ALICOCONGADC-DHISG-OCTCRTGFMGRHCEOK----- 788
 Db 1150 S-PNPNCONGATCTYLGGYSCVAGYGVNCSSEINECLSHPCQNGGTICIDLINTYKCS 1208
 QY 789 CPSTGYGCRQICD-----CLNNTG--DHITG--TCYSPGKRGARCDQAG 832
 Db 1209 CPRGTQVHCEINVDCTPFIDSTTLPEKCRNNGKCIDRVGYNCTCPGVGERCE-- 1265
 QY 833 VIIVGNLNS-LSRTSTALPADS 853
 Db 1266 ----GDVNECLSN-----PCDS 1278

RESULT 5
 A40043
 Notch protein homolog TAN-1 precursor - human
 C:Species: Homo sapiens (man)
 C:Date: 21-Apr-1992 #sequence_revision 21-Apr-1992 #text_change 02-Aug-2002
 C:Accession: A40043
 R:Ellisen, L.W.; Bird, J.; West, D.C.; Soreng, A.L.; Reynolds, T.C.; Smith, S.D.; Sklar, C.; Cell 66, 649-661, 1991
 A>Title: TAN-1, the human homolog of the Drosophila Notch gene, is broken by chromosomal
 A:Reference number: A40043; MID:91347367; PMID:1831692
 A:Accession: A40043
 A>Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra
 A:Molecule type: mRNA
 A:Residues: 1-2555 <ELN>
 A:Cross-references: GB:M73980
 C:Superfamily: notch protein; ankyrin repeat homology; EGF homology
 F:261-292/Domain: EGF homology <EGF1>
 F:994-525/Domain: EGF homology <EGF1>
 F:987-1018/Domain: EGF homology <EGF2>
 F:1149-1180/Domain: EGF homology <EGF3>
 F:1187-1218/Domain: EGF homology <EGF4>
 F:1233-1264/Domain: EGF homology <EGF5>
 F:1927-1959/Domain: ankyrin repeat homology <ANK3>
 F:1960-1992/Domain: ankyrin repeat homology <ANK1>
 F:1994-2026/Domain: ankyrin repeat homology <ANK2>
 F:2027-2059/Domain: ankyrin repeat homology <ANK4>

F:2060-2092/Domain: ankyrin repeat homology <ANK5>
 Query Match 15.3%; Score 1032; DB 2; Length 2555;
 Best Local Similarity 25.7%; Pred. No. 2,7-47;
 Matches 315; Conservative 84; Mismatches 304; Indels 522; Gaps 73;

QY 94 CCGFYSEGEKCVPHCADKCYHGR-----IAPNTQCEBGMGTGSSACDDH 143
 Db 89 CALGF--SGPLCLPDLNACLINPCRNNGTCDLLLTLEKRCRPPRGSGKSCQA----- 141
 QY 144 WGPHTSRCCCKNGALCNPIIGA--CHCAAFGRGRCE--DRECG--TYGDCNQ- 193
 Db 142 --DPCASN-PCANGQCLPFASTYICRPSFHFCTCRQDVNEGQKPRLCRHGGTCNE 198
 QY 194 ---RC-----OQNGATC--DHYTGECRCPGYTGAFC-- 222
 Db 199 VGSYRCVCRATHGTGPNCRPVVPCSPSCQNGGTCRPTGDTTHCACLPGTGNCEINI 258
 QY 223 DLCPRG--KHGPOC-----EORCP-----CQNGVCHVTVG- 251
 Db 259 DDCPRNCKNKGACVDGVNTNCPRPMTGQCTEDVDECQLMPNACQNGTCKNTHG 318
 QY 252 -ECSPSGMGTGCGO-----PCDEGRGRKNC--SQBC-- 281
 Db 319 YNCYCNVWMTGEDCSENIDDCASACFRGATCHDRVASFYCEPHGRTGLCHLNDACIS 378
 QY 282 -QCHNGGTCA--ATGQ--CHCSPYTGECQ--DECPVYGVLAETQCYNNGKCY 333
 Db 379 NPCNEGSCNDTPNVNKAICTCPSGYGPACSQVDVDESLGAN-----PCRHAGKCI 430
 QY 334 HVSA--CLCAGRAGERCEARLCPBELYIGKCKRC--PCHLENTSHPSMGE--CA 386
 Db 431 NTLSFRCQCLQGYTGPRCEIDV-----NECVSNPC--QNDATCIDQIGERCQ 477
 QY 387 CKPMSGLYC-----NE--TCSPGFYGEACQOI--CS---CQ 416
 Db 478 CMPEYBEVHCYVNTDECASSPCLNNGKCLDINFEQCECPGFTGHLCDQVDECASTPK 537
 QY 417 NGADC-----DSV--TGKCTCAPRGKI 437
 Db 538 NGAKCLDGPMTYTCVTEGYTGTHCEVDIDECDDPCHYSGCKGVATFTCLAPRYTGH 597
 QY 438 DCSF-----PCPL-----GTGLNCS-----SRGCKADAVCSP 466
 Db 598 HCFETINECSSQPCRLKGTQODPDNAYLCFLKGTTPNCEINLDCASSPCDSGTCLDK 657
 QY 467 VDG--SCYCKAGMHGVDCSIR-----CPSGTWFEGCNL-----TC----- 499
 Db 658 IDGYECACPEGYTSGMCSNINDECAGNPCNNGGTCDEGINGFTCRCEGYHDPYCLSEVN 717
 QY 500 ----OCLNGACNTLDG--TCTCAPRGNGECCEL-----P 528
 Db 718 ECNSNPVCHACRDLNGLYKCDXDPGMSGTNCIDINNNECNSNPCVNGTCKDKMTSGIVCT 777
 QY 529 CQGTLYLNAERCD-----CSHADG--HPTTGH--CRCLPGMSGVCHDSV--CAEG-- 575
 Db 778 CREFSFSPNCOITINECASNPCNLKGTCTIDVAGYKCNCLPYTGATCEVVALVAPASPC 837
 QY 576 RMPNRC-----SLPCYC-----KNGASCPDQICACAPRGNGTCQRI-----CSP 617
 Db 838 RNGEGRQSDYSEFSVCYCTPAAGKQTCVYDINECVLSCRHGASQNTNGHYRCHCQA 897
 QY 618 GFGYHRCSTCPQCVHSSGPCHH-----ITGLDCLPGFTGALCNE----- 658
 Db 898 GYSGRNCETDIDDC--RPNCHNGGSGCTDINTAFCDCLGFGFTGTCEDINECASDPCR 955
 QY 659 -----VCPSPGRGKNGAG--ICT--CTNNGTCNPIDR-----SQCTPGW 694
 Db 956 NGANCITCDVSYCTCPAGFSIHCEINTPDCTESSCFNGGTC--VDGINSFTCLCPGR 1013
 QY 695 IGSDC-----SOP-----CPRAHMGPNC--IHTCN-----CHNGAF 723
 Db 1014 TGSYCOHYVNECDSPCLLAGTCCQDGRGLHRTCPGQYTGPNQNTLVHMCDSPPCKNGK 1073

QY 699 C-----SOP-----CPAHMGPNC---IHTCN---CHNGAFC--- 724
 Db 1018 CQYVNECDSPRLHGTGTCODSYGTYKCYCPQGYTGLNCONLVKMCDSAPCKNGKRCMQF 1077
 QY 725 -SAYDECKCTPGMTGLYC---TORCPLEFY---GKDCALICO----- 760
 Db 1078 NTQY--HCECRSGMTGVCNDVLSVSCVANAQKRGIDVTLICQHGGLCVDEBGRKHYCHCQA 1135
 QY 761 -----CONGADC-DHISG-OCTCRGFMGRHCEQK----- 788
 Db 1136 GYTGSYCEDEVEDCSBNPCQNGATCTDYLGFSCKCAVAGHSGNCSSEIINECLSQPCQNG 1195
 QY 789 -----CPSGTYGCGRQICD-----CLNSTC-DHITG-TCYCS 819
 Db 1196 GTCIDLTNSYKSCSPRGTVGVCHEINVDCHPPLDPASRPRKPFNNGTCVDVGTYTC 1255
 QY 820 SPGMKARCDAQVITVGNLN 840
 Db 1256 PGFVGERCE-----GDVN 1269

RESULT 7

S1818
 notch protein homolog - rat
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 19-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 02-Aug-2002
 C:Accession: S1818
 R:Weinmaster, G.; Roberts, V.J.; Lemke, G.
 Development 113, 199-205, 1991
 A:Title: A homolog of Drosophila Notch expressed during mammalian development.
 A:Reference number: S1818; MUID:92111383; PMID:1764995
 A:Accession: S1818
 A:Molecule type: mRNA
 A:Residues: 1-2531 <MEI>
 A:Cross-references: EMBL:X57405; NID:q57634; PID:q57635
 C:Superfamily: notch protein; ankyrin repeat homology; EGF homology
 F:987-1018/Domain: EGF homology <EGF1>
 F:1025-1056/Domain: EGF homology <EGF2>
 F:1233-1264/Domain: EGF homology <EGF2>
 F:1917-1949/Domain: ankyrin repeat homology <AN1>
 F:1950-1982/Domain: ankyrin repeat homology <AN2>
 F:1984-2016/Domain: ankyrin repeat homology <AN3>
 F:2017-2049/Domain: ankyrin repeat homology <AN4>
 F:2050-2082/Domain: ankyrin repeat homology <AN5>

Query Match 15.2% Score 1024; DB 2; Length 2531;
 Best Local Similarity 25.8%; Pred. No. 7.1e-47;
 Matches 315; Conservative 83; Mismatches 286; Indels 536; Gaps 74;

Y 86 TMYRRKSGCCPFGYSEGENCV---PHCADKCVH-GRCI---APNTQCCEPFGMGTCNS- 137
 Db 121 TLTEYKCRCPGWM-SGKSCQADPCASNPNCANNGQCIFPESSTYICGCPFGHGPICRQD 178
 QY 138 -----ACDGDHMGPHC-----TSRCCKNGALCNP--- 162
 Db 179 VNECSQNGELCHNGTCHNEISYRCACNATHTPHCELPIYPCCSPSCQNGTCTCAPTD 238
 QY 163 ITGACHCAAGFRGWRCEDECEQGTGNDCHQRCQCONGATC-DHY-TGRCRCRPGYTGA 220
 Db 239 TTHEACALPFGAGNCEENV-----DCPGN-NCKNGACVDTVNTYCRCPREPTGY 291
 QY 221 C-EDLCPRGKHRCQECQRP--CONGVCHAVTG--ECSGPGMGMTYCGQ----- 266
 Db 292 CTEDV-----DEC-QLMPNACQNAAGTCHNSHGVCNVCVWGTEGDESDNIDDCASNA 343
 QY 267 -----PCDEGRFGKNC--SQEC-----OCHNGTCA--ATGQ--CHCSG 301
 Db 344 CQGGATCHDRVASFCECPHGRTGLLCHINDACISNPESSNCDNPNVNGALCTCPRG 403
 QY 302 YTGRCQ---DECPYGVGLCAETCCOCYNGSKCYHVSQA--CLCEAGFAGERCEARLCP 356
 Db 404 YTGPRACSDVDICALGAN-----PCENHAKCLNTLGSFPCQCLQGTGTRCEIDV-- 453

QY 357 EGLYIKCDKRC---PCHLENTSCHPMGSE--CAKPGWSGLYC----- 396
 Db 454 -----NECISNPC--QNDATCLDQIEFGQICIMGVEGYCEINWDECCSPCLHN 502
 QY 397 -----NE--TCSPGFGEACQ---QICS---CONGADC--DSVTCKTCAPEFGKD 438
 Db 503 GNCVDRKINEFLCQCKRGSGLHCOYDYDECASTPCKNGAKCLDGNNTYTCTVEGTGTH 562
 QY 439 CST-----PCPLGTGYNCSRCGKNDAVCSPYVDSCTCKAGW----- 477
 Db 563 CEVDIDCEDPDPCHLGL-----CK-DGVAT---FTCLCQPGYTGHHCEFTINBCH 608
 QY 478 -----HGDCSTR-----CPSGTWGECNLTQ-----CLNGACNTLDG-ITCAP 518
 Db 609 SOPCRHGGTCQDRNYYLCLCLKGTTPNCEINLDCASNPDSCGLDKIDGYEACQEP 668
 QY 519 GMRGKCEL-----PCODGTGYNCAERDCSHADCHPTT----- 554
 Db 669 GTTSGMCAVNIIDECAGSPCHNGTCEBDGAGFTC--RC---PEGHDPCLSEVNECNS 722
 QY 555 -----GHCR-----CLPGMSGVHCD-----SVCAEG 575
 Db 723 NPCIHGACRDGLNGKCDCAPEMSGTNCDINNNECESNPVNGTCMDTSGVETCRG 782
 QY 576 RMGPNC-----SLPCT-----CKNGA 591
 Db 783 FSGPNCQTNINECAASNPCINQGTCLIDVAGYKCNCLPYTGATCEVILAPCATSPCKNSG 842
 QY 592 SCSPDGI-----CEGAPGRGTQ-----RCSGPGFGH 622
 Db 843 VKESIEDYESSCVCPPTGQGTCEIDINECVKSPCRHGAQSCQNTNGSTRCLQAGYTR 902
 QY 623 RCSQTPQCVHSSGPCHH-----ITGLCDLPGFTGALCNE----- 658
 Db 903 NCESDIDC--RPNCHNGSGCTDGVNAFAFCDLPGQFAFCEEDINECATNPQNGANC 960
 QY 659 -----VCSGPRKRCAG--ICT--CTNNGTCPIR-----SCQTPRNGISGC 699
 Db 961 TDCVDSYTCCTPTGPNHGCENNTPDCTESSFCNGTGC--VDGINSFTCLCPRGFTGYSY 1018
 QY 700 -----SOP-----CPAHMGPNC---IHTCN---CHNGAFC--- 724
 Db 1019 QYDVNECDSPRLHGTGTCODSYGTYKCYCPQGYTGLNCONLVKMCDSAPCKNGKRCMQF 1078
 QY 725 SAYDECKCTPGMTGLYC---TORCPLEFY---GKDCALICO----- 760
 Db 1079 TQY--HCECRSGMTGVCNDVLSVSCVANAQKRGIDVTLICQHGGLCVDEBGRKHYCHCQA 1136
 QY 761 -----CONGADC-DHISG-OCTCRGFMGRHCEQK----- 788
 Db 1137 YTGSYCEDEVEDCSBNPCQNGATCTDYLGFSCKCAVAGHSGNCSSEIINECLSQPCQNG 1196
 QY 789 -----CPSGTYGCGRQICD-----CLNSTC-DHITG-TCYCS 820
 Db 1197 TCIDLTNTYKSCSPRGTVGVCHEINVDCHPPLDPASRPRKPFNNGTCVDVGTYTC 1256
 QY 821 PGMKARCDAQVITVGNLN 840
 Db 1257 PGFVGERCE-----GDVN 1269

RESULT 8

A49128
 cell-fate determining gene Notch2 protein - rat
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 21-Jan-1994 #sequence_revision 18-Nov-1994 #text_change 02-Aug-2002
 C:Accession: A49128
 R:Weinmaster, G.; Roberts, V.J.; Lemke, G.
 Development 116, 931-941, 1992
 A:Title: Notch2: a second mammalian Notch gene.
 A:Reference number: A49128; MUID:93202015; PMID:1295745
 A:Accession: A49128
 A>Status: preliminary; not compared with conceptual translation

A:Molecule type: mRNA
 A:Residues: 1-2471 <ME1>
 A:Experimental source: Schwann cell
 A:Note: Sequence extracted from NCBI backbone (NCBI:127811)
 C:Superfamily: notch protein; ankyrin repeat homology; EGF homology
 F:264-995/Domain: EGF homology <EGX1>
 F:799-830/Domain: EGF homology <EGF1>
 F:877-908/Domain: EGF homology <EGX2>
 F:1029-1060/Domain: EGF homology <EGF>
 F:1067-1098/Domain: EGF homology <EGX3>
 F:1153-1184/Domain: EGF homology <EGF3>
 F:1191-1222/Domain: EGF homology <EGX4>
 F:1876-1908/Domain: ankyrin repeat homology <AN1>
 F:1809-1941/Domain: ankyrin repeat homology <AN2>
 F:1943-1975/Domain: ankyrin repeat homology <AN3>
 F:1976-2008/Domain: ankyrin repeat homology <AN4>
 F:2009-2041/Domain: ankyrin repeat homology <AN5>

Query Match 14.88; Score 998; DB 2; Length 2471;
 Best Local Similarity 24.48; Pred. No. 1,7e-45;
 Matches 321; Conservative 79; Mismatches 322; Indels 596; Gaps 70;

17 CHWIGTASPLNLEDPNVCSEHVSSTVQESY-----PHPEQIYVSTCDIL 64
 121 CHMLS-----WDYECTQVGTGKCCOMTDVCLSHPCEN--GSTCSSVA 163
 65 NMFCTHRVSYRATYRNGEKTMRKRSQCCPGFEGSEMCVPRCAKCVHGRCLAPRTC 124
 164 NQFSC-----RCPAGI--TGQKCDADINEDIDGRKQHGQGTG 198
 125 QCEBGMGTNCSSACDSDHMGPHCTS-----RCQCKGALC--NPITGACHCAAGFRG 175
 199 LNLFF--GSYRQ--CPQRFTHQHDSPYPCAPSPVNGGRCRGTDGTSCHCLPREG 254
 176 WRCDRCBQGTGMDCHQRCQCCQNGATC-DHY--TEBCRCPRGYGALC-EDLCPRGKHG 232
 255 SNCRNTD-----DCPNH--KQNGGVCDVNTYNCRCRPPQMTGOFCTEDV-----D 300
 233 QC-EORPCONGVGYCHVYTG--ECSGPGMWGTVCGSP----- 267
 301 ECLLDPRNACQNGTCTNRNGGYCYCVNMGSGDSCSEIDDCARASCTPGSTCIDRVASF 360
 268 ---CPGFRGKNC--SQEC--QCHNGTCDG--ATGQ--CHCSFGYGERGQ---DECP 312
 361 SCLEPREGAGLLCHLDACISNPCHKALDPTNLNGYICTCQAYKAGACSTDVDECA 420
 313 VGTGVLCAETCCQCVNGKCHVSGA--CLCEAFAGERCE----- 351
 421 M-----ANSNPEHAGKCVNTDGAFFHCCLKYGAGPRCEMDINECHSDPCQNDATCLD 473
 352 ---ARLCPGLYGIKCDKR-----CP-----CHLE- 373
 474 KIGFTCLCMRFGKGVCELEVNECQSNPCVNGQCCVDKVRPQCLCPRTGTYCQIDTI 533
 374 ---NTHSC--HPMSEGCACKPGWSGLYCNET----- 399
 534 DDCSFTPLNCAKCIDHNGYEGCQATGFTGLCDENIDNDRPCHHGQCGDIDSYTC 593
 400 -CSGFGFEGAC-QQICSC-----QNGADCBVTG--KCTAPRGKIGDC-----STP 442
 594 ICNNGYGAICSDQIDECYSSPCLNDGRCIDLVNGYQCNCPGTSGLNCEINFDDCASNP 653
 443 CPLGTY--GIN-----SRCG-----CKNDAYC----- 464
 654 CLMGACVADGINRYSCVSPGFTGRCNIDIDECASNPCRKDATCINDVNGFRMCMPBPH 713
 465 ---SP--VDGSCT-----CKAGMHGVDCSI----- 484
 714 HPSCYQVNECLSSPCIHGNTGSLGYKCLCDAGWGINCEYVKNNECLNFCQNGGTGN 773
 485 ---KCPSGTWGFGCNLTQ-----CLNGAC----- 507
 774 NLVNGYRCTCKKGFKNYCNQJNIDECASNPLCNGTCLDVSQYTCMCLPYTGKNCQTV 833

508 -----NTLDDGTCTCAPGWRGEKELPCQDGTGTLNCAERDCSHAD 548
 834 LAPSPNPECNAAYCKEAPNESEFTCLCAPWQQRQCVDDVDE-----CVSK--PCNNNG 886
 549 GCHPPTTGH--CRCLPGNSGVHCDSDVCAEGRFNGPNCISLPCYCKNGASC--SPDDGICECAP 604
 887 ICHHTQGSYNKECPRGSGMCEE-----DINDCLANPCQNGSGCYDKKNTFSCLCLP 939
 605 GFRTGTCR-----ICSPFGYGHRCQSQTCPQCVHSS----- 635
 940 GFVQDKCQTDNNECLSEPCKNKGTCSPDVNSYTCPCPAGFHVGCENNIDECTESSCPNG 999
 636 GPCHHITGL-----DDCLPFGTALC-----NE-----VPPSRRFG 666
 1000 GTC--VQINSFSLCLPVGFTGFPCLDLINDINCSSNPCLNCGTVDGLGTACTCTPLGTYG 1057
 667 KNC---AGICT--CTNNGTC--NPIDRSCQCYEGMIGSDSQ----- 701
 1058 KNCQTLVNLCSRSPCKNKGTCQAEKARPRCLCPPGWMSAYCDVNLVSCKAALQKYPVE 1117
 702 -----PCPPAHMGPNC--HTC--NCHNAPGSAYDG--ECKCTP 735
 1118 HLCQHSGLICINAGNTHHCQCPFLGTGYSCEBQIDECASNPCQHGATGSDFTGYSRCEQVP 1177
 736 GMTGLYCGNR-----CPGFGV-----KDCALICQON 763
 1178 GYQGVNCEYEVDQCONPCQNGTCTIDLVNHFKCSCPGTRGLLCEENIDDCAPRCLN 1237
 764 GADC-DHISG-----QCTCRTEGM 781
 1238 GGQCVDRIGYGSCHCLPGRFAGERCEGDIENCLSNPCSEGLDQIKLNNQYCVCRSAFT 1297
 782 GRHCE---QKCPSTYTGICRQICDCLNNTSTCDHITGT-----CYSGFMKGARCDQA 831
 1298 GRHCETFLDVCPRK-----PCINGTCVAAVSNPDPGFIGRCPPGFGSARGCOS 1345

RESULT 9
 S42612
 Transmembrane protein precursor - zebra fish
 C:Species: Brachydanio rerio (zebra fish)
 C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 02-Aug-2002
 C:Accession: S42612
 R:Bierkamp, C.; Campos-Ortega, J.A.
 Mech. Dev. 43, 87-100, 1993
 A:Title: A zebrafish homologue of the Drosophila neurogenic gene Notch and its patte:
 A:Reference number: S42612; MUID:94128602; PMID:8297791
 A:Accession: S42612
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-2437 <BIE>
 A:Cross-references: EMBL:X69088; NID:9433866; PID:CAA48831.1; PID:9433867
 C:Superfamily: notch protein; ankyrin repeat homology; EGF homology
 F:755-786/Domain: EGF homology <EGF1>
 F:1023-1054/Domain: EGF homology <EGF2>
 F:1185-1216/Domain: EGF homology <EGF3>
 F:1315-1947/Domain: ankyrin repeat homology <AN1>
 F:1948-1980/Domain: ankyrin repeat homology <AN2>
 F:1982-2014/Domain: ankyrin repeat homology <AN3>
 F:2015-2047/Domain: ankyrin repeat homology <AN4>
 F:2048-2080/Domain: ankyrin repeat homology <AN5>

Query Match 14.68; Score 987; DB 2; Length 2437;
 Best Local Similarity 24.88; Pred. No. 6.3e-45;
 Matches 310; Conservative 81; Mismatches 320; Indels 538; Gaps 70;

91 KSQCCPGFYSBEGMCVHCHADKCVHRC-----IAPNTQCCPRMGCTCSSACD 140
 85 KCDCVIGF--SDRLCLPRVNHACNNSPCRNKGCTSLTLDTFTFCRCPGSGKTCQLA-- 140
 141 GDHWPCTSRCCQKNALCNPTIG--ACHCAAGFRMGCEDRCEQCTYGNDCH-QRCQC 197

```

Db 141 -----DPCASN-PCANGGCSAFESHYICTCPNFHGQTCRODY-----NECAVSPSPC 188
QY 198 QNGATCDHTYGE--CRCPGTYGAFCEDL-----CPSPK 229
Db 189 RRGGTCINVGSTLCRCPPETGPHQRLXPCLPSPCSGGTCVTSPTHTCSLPPF 248
QY 230 HGPQCE-----ORC-----PCQNG 243
Db 249 TQOTCEHNVDCTQHACENGSPCIDINTYNCHDKHMTGOVCTEDVDECELSPNACQNG 308
QY 244 GYCHHTG--ECSBPGMNGTYCGO-----PCEGFRGN 276
Db 309 GTCCHNTGGFHCYCVANGWDSCSENIIDCASAACHGATCDHRVASFECEPHGTGL 368
QY 277 C--SQEC--OCHNGSTCA--ATGO--CHSPGYTGRCQ--DECPYGTGVLCAETC 324
Db 369 CHLDACISNPOCKGKNCJTNVSGAICTCPGTYGSCANDIDECISGAN----- 420
QY 325 QCVNGKCYHVSQA--CLCEAGFAGERCEARLCPEGLYKICKDKRCPCHEMTHSCHPMS 382
Db 421 PCEHGGRCINTKGSFQCKCLQGEYGRCEMDV-----NEC-KSNPC--QNDATCLDQI 470
QY 383 G--ECACKRSGMGLVCNER-----CSRFYGEACQ--QIC 413
Db 471 GGFHCICMFGYBGFVFCQINSDDCASOPCLNGKCIDKINSFHCPCRGFSGLQVDVDEC 530
QY 414 S--CONGADCDVYTK--CTCAPGFKGIDC-----STPCPL-----GYGINSR 455
Db 531 ASTPCKNAGKCTDGPKNKYCECTPGFSGHCELDINECASPSCHVCHDYAASFCDR 590
QY 456 CG-----CKDANVCSPIVDS--CTCKAGMHGVDCSIR----- 485
Db 591 PGTGRLCETNINECLSOPCRNGGTQDRENAVICTCPRTGTGVNCEINIDCKRRPCDY 650
QY 486 -----CPSGTGFSCNLTQO-----CLNGACNTLDG-----TTCARGMG 522
Db 651 GKCIDKINGEYCEYSGSMCNINIDCALNPGHNGTIC--IDGVNFTCLCPRGFED 708
QY 523 ERC-----ELPCQDGYTLNC--AERC--DCSHADGCHP 552
Db 709 ATCLSQHNECSSNPICHSGLDQINSYRCVCEAGWMGRMCDININCLSNPCVNGSTCD 768
QY 553 -RTGH--CRCLPGMSGVHCDVCAEGRMGP-----NCSL----- 583
Db 769 MTSGLCTCRAGFSGPNCCMINIECASNPCLNMGSCIDVAGFKCMPLPYGEVENVL 828
QY 584 -PCY--CANGASCSPD-----GICE----- 601
Db 829 ABCSPRCKNGVCHRESDFQSFSCNCPAGMGTCEVDINECVRRPCTNGVCEMLRGG 888
QY 602 ---CAPGRTTCOR-----ICSPGYGHRCSQTCPCV 632
Db 889 FQCRPNPFTGALCENDIDCEPNPCSNGVQCDRVNGFVCLAGFRGERCAEDIDECV 948
QY 633 HSSGPHHTITGLCDLPFTGALCNEVCSGRFKNC--AATCT--CTNNGTCNPIR 686
Db 949 --SABCRKNGNCTDCVNST--CS--CPAGFSGINCEINTPDCIESSEFNNGTC--VGG 999
QY 687 ---SCOCYPGWIGSDC-----SQP-----CPRHMPNC--IH 714
Db 1000 ISSFCVCLPFTGNYCQMDVNECDSPQNGSGQDGYTKCICPHRYTGLNCSLYR 1059
QY 715 TCN---CHNGAFC--SAVDECKCTPGWTGLYCTO----- 744
Db 1060 WDDSPCKNGSGCMQOGASFTQCSAGWTGYCDVPSVCEVAARQGVAVLCHRAQ 1119
QY 745 -----RCPLGFYKDC--ALIQ--CONGADC-DHISG--QCTCRGFPMGRHE 786
Db 1120 CVDAGNTHLCRCQAGYTSYCEQVDECCPNPCQNGATCTDYLGGSCBVCVGHMNS 1179
QY 787 OK-----CPSGTYGRCQICD-----CLNN 807
Db 1180 KEINCLSLQPCQNGGTCLDVNTYTKCSPCRGTGIVHCEIDIDDCSPSVDPLTGEPRCENG 1239

```

```

QY 808 SPC-DHITG--TCYSPGMRKARODQAVITVGNLSRSTALPADSY 854
Db 1240 GRCVDRVGGYGCVCAPGFERCE-----GDVNE-CLSDPCDPSGST 1280

```

RESULT 10

```

A24420
notch protein - fruit fly (Drosophila melanogaster)
M:Alternate names: neurogenic repetitive locus protein
C:Species: Drosophila melanogaster
C:Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text change 10-Sep-1999
C:Accession: A24420; A24768; S09358; A05267
R:Kidd, S.; Kelley, M.R.; Young, M.W.
Mol. Cell. Biol. 6: 3094-3108, 1986
A:Reference number: A24420; MUID:87064624; PMID:3097517
A:Accession: A24420
A:Molecule type: DNA
A:Residues: 1-2703 <KID>
A:Cross-references: GB:K03508; NID:9157991; PIDN:AAA28725.1; PID:9157993
R:Wharton, K.A.; Johansen, K.M.; Xu, T.; Artavanis-Tsakonas, S.
Cell 43, 567-581, 1985
A:Reference number: A24768; MUID:86079539; PMID:3935325
A:Accession: A24768
A:Molecule type: mRNA
A:Residues: 1-48, 'I', '50-118, 'R', '120-230, 'I', '232-256, 'N', '258-266, 'A', '268-872, 'R', '874-9
A>Note: the authors translated the codon ATC for residue 49 as Thr, ATT for residue 2
R:Antz, D.
Nucleic Acids Res. 17, 6463-6471, 1989
A:Title: Hypervariability of simple sequences as a general source for polymorphic DNA
A:Reference number: S09358; MUID:89385974; PMID:2780284
A:Accession: S09358
A:Molecule type: DNA
A:Residues: 2505-2551, '0000', '2552-2576, 'E', '2578-2604 <TAU>
R:Wharton, K.A.; Vedyobnick, B.; Finnerty, V.G.; Artavanis-Tsakonas, S.
Cell 40, 55-62, 1985
A:Title: opa: a novel family of transcribed repeats shared by the Notch locus and oth
A:Reference number: A05267; MUID:85099329; PMID:2981631
A:Accession: A05267
A:Molecule type: DNA
A:Residues: 2504-2576, 'E', '2578-2611 <WHA>
C:Genetics:
A:Gene: notch; opa
A:Cross-references: FlyBase:FBgn0004647
A:Map position: 8.96-9.36
A:Introns: 53/3; 84/3; 171/3; 240/3; 283/3; 2333/3; 2436/3; 2588/3
C:Superfamily: notch protein; ankyrin repeat homology; EGF homology
C:Keywords: differentiation; tandem repeat; transmembrane protein
F:27-43/Domain: transmembrane #status predicted <TM1>
F:297-328/Domain: EGF homology <EGF1>
F:530-561/Domain: EGF homology <EGF1>
F:568-599/Domain: EGF homology <EGF>
F:988-1019/Domain: EGF homology <EGX2>
F:1064-1095/Domain: EGF homology <EGX3>
F:1187-1218/Domain: EGF homology <EGX3>
F:1746-1762/Domain: transmembrane #status predicted <TM2>
F:1983-1982/Domain: ankyrin repeat homology <AN1>
F:1983-2015/Domain: ankyrin repeat homology <AN2>
F:1988-2004/Domain: transmembrane #status predicted <TM3>
F:2017-2049/Domain: ankyrin repeat homology <AN3>
F:2050-2082/Domain: ankyrin repeat homology <AN4>
F:2083-2115/Domain: ankyrin repeat homology <AN5>
F:2538-2568/Region: glutamine-rich
F:2538-2568/Domain: neurogenic repetitive element #status predicted <OPA>

```

Query Match

```

14.5%; Score 978.5; DB 1; Length 2703;
Best Local Similarity 26.8%; Pred. No. 2e-44;
Matches 290; Conservative 102; Mismatches 297; Indels 395; Gaps 70;

```

```

QY 7 SCL-----SPICLLCHWIGTASPLNLE--PNVCSHWSYSTVQESYHPDPQIYTC 60
Db 502 SCLDPRGFRRCVCMGFTGTQCEIDIDECQSNPC-----LNDGTC 541

```



```

OY 484 -----IRCPSSGWRGNCNLTG-QCLNG-----GACTTLDG-----TCTCAPGWRNG 522
Db 596 GSKCLDLYDKYLCRCRPPGTGVCNENIDDCASNPTCFVCR--DGINRYDCVCOGPGFTG 652
OY 523 ENCEL-----PCODGTGTLNLC-----AERCDCSHADGCH 551
Db 653 PLCAINEINCASSPCGEGSSCVDSGNGRFLCPLPPSLPLCLPANHPCHAKRCSHG-VCH 711
OY 552 PTTG--HCRCLPFGMSGVHCDVCAEGRWGPNCSLPCYCKNGASCPDDGI---CECAPGF 606
Db 712 DAGGFRVCPEPWSGPRCSQSLA-----PDACESQPCOAGGTC--SDIGFRVCAPGF 765
OY 607 RGTTCORI-----CSPFGYGHNCSTQCPQCVSS-----G 636
Db 766 GGHCEVSLPCTPSLCEHGHCESPDRLLTVCSCPPGMOGPRCQDDVDEACAGASPCGPRG 825
OY 637 PCHNITG--LDCDLPFGFALCNE-----VCPSGRFGKNC 669
Db 826 TCTNLPGNPRCICHGTYGPRPCDDDDIDODPNRCLHGSSCDGVGSFSCSLDGFAGPRC 885
OY 670 A-----GICF-----CTNNGTCNPIDR 686
Db 886 ARVDDECLSSPCGPCTCTDHVASFYCACPPGYGFHCEIDLPCDSCPSSCFNCGTC--VDG 943
OY 687 ---SCCYPGWSDC-----SOP-----CPRAHMGPNCIHTCN----- 717
Db 944 VSSFSCLCRPGYTGTHCYEADPCRSRPLCHGICNPHTPHEC--TCREGTSGQCONP 1001
OY 718 ---CHNGAFCSAYDGECKTPGWTG-----LYCTOR----- 745
Db 1002 VDMCSQAPCQNGRCVYTGAYCICPPGWSRLCDIQLPCTEAAAMQNGVRLEQLQEGSK 1061
OY 746 ---CPGLFYKDC-----ALICQCONGACCDHISG--QCTCRTEGMRICE 786
Db 1062 CIDGRSHYCVCPESRTSHCEHEVDPCTAQCQHGTCRGYMGGYVCECPAGYAGDSCE 1121
OY 787 Q-----KCPSGTYGK-----KQICD-----CLNNS 808
Db 1122 DNIDECASQPCQNGSSCIDLVARYLCSPPGTGLVLCIEINEDCDLGFSLDSGVQCLHNG 1181
OY 809 TCDHITG--TCYSPGKMGARCD 829
Db 1182 TCVDLVGGFRNCNCPGTYGLHCE 1204

```

RESULT 12

notch3 protein - human

C:Species: Homo sapiens (man)

C:Date: 24-Jul-1998 #sequence_rev150n 24-Jul-1998 #text_change 02-Aug-2002

C:Accession: S78549; S71825

R:Joutel, A.; Tournier-Lasserre, E.

submitted to the EMBL Data Library, April 1997

A:Reference number: S78549

A:Accession: S78549

A:Molecule type: mRNA

A:Residues: 1-2321 <J001>

A:Cross-references: EMBL:U97669; MID:g2668591; PIDN:AB91371.1; PID:g2668592

R:Joutel, A.; Corpechot, C.; Ducros, A.; Vahedi, K.; Chabrial, H.; Mouton, P.; Alamowicz

X, M.M.; Weissbach, J.; Bach, J.F.; Bousser, M.G.; Tournier-Lasserre, E.

A:Title: Notch3 mutations in CADASIL, a hereditary adult-onset condition causing stroke

A:Reference number: S71825; MID:97032728; PMID:8878478

A:Accession: S71825

A:Status: nucleic acid sequence not shown

A:Molecule type: DNA

A:Residues: 67-113;138-194;268-333;G',335-346;536-613;716-765;1240-1279;1815-1888 <J002

C:Cross-references: EMBL:U97669

C:Genetics:

A:Gene: notch3

A:Map position: 19p13.1

C:Function:

A:Description: may be involved in pathogenesis of CADASIL, causing a type of stroke and

```

C:Superfamily: notch protein; ankyrin repeat homology; EGF homology; laminin-type EGF
C:Keywords: tandem repeat; transmembrane protein
F:123-155/Domain: EGF homology <EGX1>
F:162-194/Domain: EGF homology <EGX1>
F:240-271/Domain: EGF homology <EGX2>
F:318-349/Domain: EGF homology <EGX>
F:473-504/Domain: EGF homology <EGX3>
F:853-884/Domain: EGF homology <EGX3>
F:928-959/Domain: EGF homology <EGX4>
F:1070-1126/Domain: laminin-type EGF-like homology <LEG>
F:1838-1870/Domain: ankyrin repeat homology <AN1>
F:1871-1903/Domain: ankyrin repeat homology <AN2>
F:1905-1937/Domain: ankyrin repeat homology <AN3>
F:1938-1970/Domain: ankyrin repeat homology <AN4>
F:1971-2003/Domain: ankyrin repeat homology <AN5>

Query Match 14.4%; Score 974; DB 2; Length 2321;
Best Local Similarity 25.3%; Pred. No. 3e-44;
Matches 304; Conservative 69; Mismatches 313; Indels 514; Gaps 68;

OY 107 PHCAD--KCVH-GRCI-APN---TQCEPFGNGTNCSSACDGDHMGPRCTRCCKNGAL 159
Db 41 PRCIDGSPCANRGKRTQLPSPREACLPFGWVGERQLE-DPCHSGP-CAGRGVQSSV 98
OY 160 CNPLTGACHCAGFERGWRCE--DRCEQGTGNDCHQRCQCONGATYCDHVTGE-----CRC 212
Db 99 AGTARFSCRCRPGFRGPRCPSLPDCLSS-----PCAHGANCs--VBPDRFLCSC 146
OY 213 PRGTGATGCE---DLCPFGKHGPRQCEQRCPRONGVCHHTVG--ECSPSGMMGTVCQP 267
Db 147 PRGTGATGCRSDVDCEKRGV-----PCRHGCTLNTGTSFRQCQPAATYGPLCENP 197
OY 268 CPEGRFGKNCSEQCHNGTCDAA--TGQCHSPGYTGSRQ--DECP----- 312
Db 198 AVP-----CAPS-PCRRGGRGSGDLTYACALPREGGNCENYNDCCGHNCINGT 250
OY 313 ---VGTGYVLC-----AETCQ-----CVNGKCYHVS--ACICEAGFAGER 349
Db 251 CVDGNTVNYCCCPPEMTGQFCTEDVDECOLPNMACHNGTGFNTLGHSHCVANGWITS 310
OY 350 CEARL-----CPGGLYGIKC--DKRC--PCLEMTHTSHCP 380
Db 311 CSQNDICATVACFRGATCDHVAIFYCACPMKGTGLCHLDACVSNPCHEDIAICDTNP 370
OY 381 MSGE--CAKPRGWSGLVCNE----- 398
Db 371 VNGRAICTCPGPGFTGAGCDDQDVDECSIGANPCHELRGVNTQSGFLCQGGTYGPRCT 430
OY 399 -----TSPGYGERACQ-----QICSONGADC-DSVTG 426
Db 431 DVNECLSGPCRNQATCLDRIGQFTICMAGFTGYCEVDIDECOSSPCVNGVCKDRVNG 490
OY 427 -KTCAPGFKGIDC-----STPCPLGTV-----GINGSRCG-----C-KNDAY 463
Db 491 FSCCTPSPGFSSTQLDVDECASTPCRNAGKACVQDPGDEY--RCABGFSTLLCDRNVDD 548
OY 464 CSP-----VDG-----SCTCKAGMHGVDCS----- 527
Db 549 CSPDPCRHRCVDDIASFSACAGYGTGTCEQSVDECRSPCRHGKCLDLYDKYLCRC 608
OY 487 PSCTGWSGCHNLTG-QCLNG-----GACTTLDG-----TCTCAPGWRNGECEL----- 527
Db 609 PSCTGWSGCHNLTG-QCLNG-----GACTTLDG-----TCTCAPGWRNGECEL----- 527
OY 528 ---PCODGTGTLNLC-----AERC-----DCSHADGCHPTTG--HCRCLPFGWS 564
Db 667 GEGGSCVDGENGFRCLCPPOSPLPLCLPSPHPCNHEPCSHGT-CYDAPGGRVCCEGWS 725
OY 565 GVHDS-----VCABGRWGPNCSTL--PCY-----CKNGAS 592
Db 726 GPRCSQSLADACESQPCRAAGTCCSSDGMGFHCTCPRGVGRGRCCELLSPCTPNCHEGGR 785
OY 593 CSPDDG---ICECAPGFRGTTGCR-----ICSPGFGYGRH 623

```


A:Cross-references: GB:X17530; NID:g10225; PID:g667061
 A:Accession: B40136
 A:Status: preliminary; not compared with conceptual translation
 A:Molecule type: DNA
 A:Residues: 181-251,329-370,'R',372-408,'RA',411-441<DE>
 A:Accession: C40136
 A:Status: preliminary; not compared with conceptual translation
 A:Molecule type: DNA
 A:Residues: 'K',747-821,898-978<DE>
 R:Hursh, D.A.; Andrews, M.E.; Raff, R.A.
 Science 237, 1487-1490, 1987
 A:Title: A sea urchin gene encodes a polypeptide homologous to epidermal growth factor.
 A:Reference number: A23116; MUID:87319677; PMID:3498216
 A:Accession: A23116
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 'S',280-481,786-1064<HUB>
 A:Cross-references: GB:M17421; NID:g161474; PIDN:AAA30050.1; PID:g552260
 R:Hunt, L.T.; Barker, W.C.
 FASEB J. 3, 1760-1764, 1989
 A:Title: Avidin-like domain in an epidermal growth factor homolog from a sea urchin.
 A:Reference number: A43131; MUID:89196806; PMID:2784773
 A:Contents: annotation
 C:Comment: EGF homology repeats 10-17 are spliced out in the short form (fibropellin lb)
 C:Superfamily: C1r/C1s repeat homology; EGF homology
 F:1-19/Domains: signal sequence #status predicted <Sig>
 F:10-1064/Product: fibropellin I #status predicted <FIB>
 F:23-54/Domains: EGF homology <EG01>
 F:57-175/Domains: C1r/C1s repeat homology <C1R>
 F:180-211/Domains: EGF homology <EG02>
 F:218-249/Domains: EGF homology <EG03>
 F:256-287/Domains: EGF homology <EG04>
 F:294-325/Domains: EGF homology <EG05>
 F:332-363/Domains: EGF homology <EG06>
 F:370-401/Domains: EGF homology <EG07>
 F:408-439/Domains: EGF homology <EG08>
 F:446-477/Domains: EGF homology <EG09>
 F:484-515/Domains: EGF homology <EG10>
 F:522-553/Domains: EGF homology <EG11>
 F:560-591/Domains: EGF homology <EG12>
 F:598-629/Domains: EGF homology <EG13>
 F:636-667/Domains: EGF homology <EG14>
 F:674-705/Domains: EGF homology <EG15>
 F:712-743/Domains: EGF homology <EG16>
 F:750-781/Domains: EGF homology <EG17>
 F:788-819/Domains: EGF homology <EG18>
 F:826-857/Domains: EGF homology <EG19>
 F:864-895/Domains: EGF homology <EG20>
 F:902-933/Domains: EGF homology <EG21>
 F:936-1064/Region: Avidin-like
 F:83-34,28-43,45-54,62-88,180-191,185-200,202-211,218-229,223-238,240-249,256-267,261-27
 57,453-466,468-477,484-495/Duplicate bonds: #status predicted
 F:489-504,506-515,522-533,527-542,544-553,560-571,565-580,582-591,598-609,603-618,620-62
 08,810-819,826-837,831-846,848-857,864-875,869-884,886-895,902-913,907-922,924-933/Dupl

Query Match 14.2%; Score 954.5; DB 2; Length 1064;
 Best Local Similarity 28.0%; Pred. No. 1.6e-43;
 Matches 290; Conservative 93; Mismatches 305; Indels 347; Gaps 69;

QY 30 DPVWGSMEYSTVQESYHPPDQIYVSTCDILNFKTRIRVSRTRVRRGKEMR 89
 DB 181 DPNLCONG-----AACTDLVNDYACT----- 201
 QY 90 RKSQCPGEYSESGEMC---VPHCA-DKCVHRCIAPN---TCQCEPGMGTCSSACDG 141
 DB 202 -----CPPGF--TGRCNEIDIDECASDPQCGACVGVNVCNCVPGFSGDECCENIN- 254
 QY 142 DHMGPHTSCQCKNGALCNPIRGA---CHCAGFGWACE--DRCEGGYTGNDCHOR 194
 DB 255 -----BCAS-SPLNGGIC--VDGVNMFECTCLAGFTGVCENVIDECAS----- 296
 QY 195 CQGNKATC-DHYTG-ECRPPRYTGAFCEDDLPRGKHGQCRQRCPCQNGVCHVYTG 252
 DB 195 CQGNKATC-DHYTG-ECRPPRYTGAFCEDDLPRGKHGQCRQRCPCQNGVCHVYTG 252

DB 297 APCONGICIDINGTYCSCPLFGSGDNEN-----NDECCS-IPCLNGTCVDLVNA 349
 QY 253 --CSPSGMWGTVCGGPPCPGREGKNCOSQCCQCHNGTC-DATG-QCHSPGYTERQ 308
 DB 350 YMCVCAPGWTGPPCADNIDE-----CA-SAPONGVCEIDYNGVTCMDQPPYTGTHCE 402
 QY 309 ---DECPVGTGYLCAETCCCVNGKCYH-VSG-ACLCEAGFAGECEARLCEPGLYGR 363
 DB 403 TDIDEC-----ARPPCQNGGDCVDGVNVCICAGFGLNDE----- 440
 QY 364 CDRRCCHLENTISC--HPMSGECACKPGHSGLYCNETGSPGYGACQ--QICS---C 415
 DB 441 -----NNIDECASRPCONGAVCVGVNGFVC--TCSAGTGVLCEDINEDECSMPC 489
 QY 416 QNGADC-DSVTGK-CCAPGFKGIDSTPCPLGTGYINGCSRCKNDAYCS-PVGG-SC 471
 DB 490 LINGVCTDLVNGTYCTCAGFEGTNCETDID-----ECAS-FPCQNGARCTQVNGVYC 542
 QY 472 TCRAHGHVDC-----SIRCPG-----TWGFCNL-----TCQ----- 500
 DB 543 TCVPGYTGVLCEDINECASFPCLNGTCDNDVNGVYCAQDTSVSTCTEDRDECASAP 602
 QY 501 CLNGAC-NTLDG-TCTCAGMGRKCEL-----PCQDGYGLNCAERCCSHADGC 550
 DB 603 CLNGACMDVNVNFCVCTCLPGWEGTNCETDIDECASSPCNG--GL-CVDQVN-STY--- 655
 QY 551 HPTGRCRLPGMSGVHCDVSCAEGHMGPNCSLPCYCKNGASC--SPDDGICBCAGFRG 608
 DB 656 -----CFCLPFGTGHGCEIDECASSP-----CLNGGQCIDRDVSEVCACAGYTA 702
 QY 609 TTQO-----RISGPFYGHRCSTCPCQVHSSGPC----- 638
 DB 703 VRQINIDECASAPCONGVGVGVNVCNCAPGYTGDNCEIDEBC--ASMPCLNGA 760
 QY 639 --HHNG-LGDCPLGFGALC-----NEVPSGFGKNCAGICTCTNGTCNPIDRS 687
 DB 761 CIEMVNGYTCCVAGTGVCEIDIDEBCASAPONG-----GVCTDITNGTY----- 807
 QY 688 CQCPGWIGSDCSQPCPPAHMGFNCIHTCNCNGAFCSAYDG--ECKCTPGMTGLYCP 743
 DB 808 CACVPFGTSGNCTENIDECASDP-----CLNGGIC--VDGVNFGVCCQPPHYSGTYCE 858
 QY 744 QRCPLGFGKDCALICQONGACDHISSQ--CTCRGFGRCCE--QKCPGTYGYGC 798
 DB 859 -----ISLDACRSMPCQNGATCVNADYVECPGVAGQCEIDINECAS----- 904
 QY 799 RQICDCLNNTG-CDHITG-TCYCSPGMKARCOAGYIIVGNL-----NSLSRTSTA 848
 DB 905 ---LPCONGGLCIDGAGTTCQCRILGTYGVNCEVFGCDLGMVMECNDQVITRTSTG 961
 QY 849 LPADSYQIGAIAGIILVVLFLALFTIYRRHKOKRESSMPAVYTPAMRYVNDYTI 908
 DB 962 M-----MLGDVMTYNNRAGLYAPVYVGVASN-----NNDPFS 995
 QY 909 SG-TLPHSGNGANS 922
 DB 996 FGFTVVRDNGQSTTS 1010

RESULT 15
 T09059
 notch4 - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 11-Jan-2000
 C:Accession: T09059
 R:Rowen, L.; Mahaitas, G.; Qin, S.; Ahearn, M.E.; Dankers, C.; Loretz, C.;
 submitted to the EMBL data library, October 1997
 A:Description: Sequence of the mouse major histocompatibility locus class III region.
 A:Reference number: Z16543
 A:Accession: T09059
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1964 <ROW>

A:Cross-references: EMBL:AF030001; NID:g2564945; PID:g2564947

C:Genetics:

A:Gene: notch4

A:Map position: 17

A:Intons: 22/1; 148/1; 264/1; 305/1; 384/1; 436/1; 501/1; 539/1; 577/1; 618/1; 671679/3; 1729/1; 1761/3

C:Superfamily: unassigned EGF-related proteins; EGF homology

C:Keywords: receptor; signal transduction

F:514-545/Domain: EGF homology <EGF>

Query Match

Best Local Similarity 14.1%; Score 952.5; DB 2; Length 1964;

Matches 298; Conservative 68; Mismatches 315; Indels 445; Gaps 65;

```

QY 95 CP-GEFSEEMCVPHCADCV-----HRCIAPNT-----CQCEPBGWGTNCSACDGH 143
DB 102 CDSGF--TGDRQOTLLEELCPSPFSCNGGHCYVQASGRPOCSEPEMTGEOQLR----- 154
QY 144 MGPHTSRQCKNGALCNPTG--ACHAAGFRGNCE-----DRCEGTGYGNDG 191
DB 155 --DFCSAN--PCANGGCLATPTYPQICRCRPFEGHTECEADINECFLEBPGRPGT---SC 208
QY 192 HQ-----RCQ-----CONGATCD-----HVTGE--CRCPGTYGA 219
DB 209 HMTLGSYQCLCPVGEGRPOCKLTKGACPPGSCINGGTQQLVPEGHSTFHLCLCPGFTGL 268
QY 220 FCE-----DLCPGKHG-----PCGEORCP--CONGV 245
DB 269 DEEMNDDCVRHOCNGATCLDGLDITYTCLCPRTWKGMDCSEIDCEARPPRCRNGGT 328
QY 246 CHVITG--ECSPGSMWMTVCGP-----CPEGRFGKNC 278
DB 329 CONTAGSFHCYCVSGWGGAGCEENLDDCAATCABGSTCIDRVGSTRCLCPRGRTGLCH 388
QY 279 QE--C--OCHNGTC--DAATGQ--CHCSPGYTGERQ--DECPVGTGYVCAETCQC 326
DB 389 LEDMCISOPCHVNAOCSTNPLTGLTICQPGYSGSTCHQDLDECQMAQGG-----PSPC 443
QY 327 VNGGKCYHSGA--CLCEGAFGERCEAR-----LCPEGIX 360
DB 444 EHGGSCTINPGSFNCLCPGVTGSRCEADHNECLSQPCHPGSTCLDLATFHLCLCPPGLE 503
QY 361 GIKCD---KRC---PCHLEHTSHCHPMG--ECACKPGMSGLYCNE----- 398
DB 504 GRICEVEVNECTSNPC--LNQAAHDLNGFOCLCPGFTGARCEKDMDECSSTPCANG 561
QY 399 -----TCSPEFYGEACQICS-----CONGADCDVYTK--CTCAPGFKGIDC 439
DB 562 RCRDQPAFYCECLPGEFGEHCEKEVDECLSDPCPYGASCLDLPGAFFCLCRPGFTGLC 621
QY 440 STP-----CPLTGYI-----NCSSRCG----- 457
DB 622 EYPLCTPNMCPGQOCQGEHRAFCCLCPDGSPPCVPAEDNCPCHHGCQRSILCVDEGWT 681
QY 458 -----CKNDVCSPYDG--SCTCKAGMHGVDS--TRCPSGTW----- 491
DB 682 GPBCETELGGCISTPCAHGIGTCHPQPSGYNCTCPAGYMLTCSBEVTACHSGPCLNGSSC 741
QY 492 -----GFGCN-----LTQCLNGGACNTLDGT--CTCAPGWRGKCE- 526
DB 742 SIRPEGYCTCLPSHTGRHCQTAVDHCVASCLNGGTGVNKKPGLCATGFOGLHCEE 801
QY 527 -----LPCQGTGYLNC-----AERCDSSHADGCHPT-- 553
DB 802 KTNPSCADSPCKNKATCQDTPRGARCLSPGYTGSSCQTLIDLCAKPCPHPARCLQSGP 861
QY 554 TGHCRCLPMSGVHCD--SVCAEGRWGPNCSTLPCYCKNGASCSPPDG--ICECAPFRG 608
DB 862 SFQCLCLQGWMTALCDFTLPSCKAKMSQGIETISGLCQNGGLCT-DTGSSTFCRCRPFPG 920
QY 609 TTCQRICSPGYGHRCSQTCPOCVHSSGPCHHITGLCDCLPFTGALCNEVCPSGRFRGN 668
DB 921 KLCODNVNP-----C--EPNCHHGS--TCVPQPSGYVCO--CAPGYEGGN 960

```

```

QY 669 CAGI---C---TCTNNGTC--NPIDRSQCQYPGWIGSDC-----SQPCPPAHMGRNC 712
DB 961 CSKVLDAQCQSPCHNHGTSTSRPGFGFCACPPGFVGLRCEGDVDECLDRPHPS----- 1014
QY 713 IHTCNCHNGAFCSAYDGECKTPTGWTGLYC-----TQ 744
DB 1015 -GTACH--SLNAF--TYQCLPGHTGQRCVEYEMDLQSQPCSGNGSCETITGPPPGFTC 1069
QY 745 RCPLEFYGRDC--ALIC--OCONGADC-----DHISQCTCRTGFMGRHC--EOKCPSG 792
DB 1070 HCPKGFEPGPTCSHKALSCGIIHHCHNGGLCLPSPKPSPLCACLSGFGGPDCLTPAP- 1128
QY 793 TYGCGRQICDCLNSTCHITG-----TCYCSPPGKARCDQAG 832
DB 1129 ---GCGPPSPCLHNGTCTETPEGLNPGFQCTCPDPSPPRCQRP 1170

```

Search completed: May 9, 2003, 11:58:26
Job time: 64.0243 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 9, 2003, 11:54:50 ; Search time 122.851 Seconds

(without alignments)
853.959 Million cell updates/sec

Title: US-10-092-390-2

Perfect score: 6744
Sequence: 1 MVISINSLSTFICLLCHWT.....SSPKQEDSGSSSSSSSS 1140

Scoring table:

Gapop 10.0, Gapext 0.5

Searched: 349150 seqs, 92025710 residues

Total number of hits satisfying chosen parameters: 349150

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications_AA:*
1: /cgn2_6/ptodata/2/pubppaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubppaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubppaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubppaa/US07_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubppaa/US07_PUBCOMB.pep:*
6: /cgn2_6/ptodata/2/pubppaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubppaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/2/pubppaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubppaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/2/pubppaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubppaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/2/pubppaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubppaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/2/pubppaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6744	100.0	1140	US-10-092-390-2	Sequence 2, Appl1
2	3601	53.4	586	US-10-092-390-4	Sequence 4, Appl1
3	2506.5	37.2	1050	US-09-796-753-114	Sequence 114, App
4	1909	28.3	636	US-09-796-753-100	Sequence 100, App
5	1909	28.3	636	US-09-796-753-124	Sequence 124, App
6	1192	17.7	474	US-10-084-994-11	Sequence 11, Appl1
7	1034.5	15.3	2444	US-09-944-849-2	Sequence 2, Appl1
8	1011	15.0	241	US-10-084-994-8	Sequence 8, Appl1
9	961.5	14.3	497	US-09-796-753-116	Sequence 116, App
10	870.5	12.9	830	US-09-870-759-134	Sequence 134, App
11	870.5	12.9	830	US-09-870-759-140	Sequence 140, App
12	870.5	12.9	830	US-09-842-758-57	Sequence 57, Appl1
13	815	12.1	639	US-09-900-449A-4	Sequence 4, Appl1
14	808	12.0	865	US-09-842-758-20	Sequence 20, Appl1
15	795.5	11.8	601	US-09-900-449A-5	Sequence 5, Appl1
16	790	11.7	1055	US-10-219-248-2	Sequence 2, Appl1
17	790	11.7	1055	US-10-219-247-2	Sequence 2, Appl1
18	790	11.7	1055	US-09-855-722-2	Sequence 2, Appl1
19	789	11.7	1212	US-10-219-248-3	Sequence 3, Appl1

20	789	11.7	1212	9	US-10-219-247-3	Sequence 3, Appl1
21	789	11.7	1212	10	US-09-855-722-3	Sequence 3, Appl1
22	789	11.7	1238	10	US-09-944-849-4	Sequence 18, Appl1
23	783.5	11.6	934	9	US-09-842-758-18	Sequence 18, Appl1
24	779	11.6	296	9	US-09-866-050A-458	Sequence 458, App
25	779	11.6	299	9	US-09-866-050A-192	Sequence 192, App
26	779	11.6	299	9	US-09-866-050A-332	Sequence 332, App
27	779	11.6	1238	9	US-10-219-248-5	Sequence 5, Appl1
28	779	11.6	1238	9	US-10-219-247-5	Sequence 5, Appl1
29	779	11.6	1238	10	US-09-855-722-5	Sequence 5, Appl1
30	769	11.4	2201	12	US-10-100-912-2	Sequence 2, Appl1
31	768.5	11.4	572	9	US-09-900-449A-7	Sequence 61, Appl1
32	766	11.4	2167	10	US-09-778-927A-61	Sequence 6, Appl1
33	757	11.2	1036	10	US-09-995-593A-6	Sequence 7, Appl1
34	757	11.2	1187	10	US-09-995-593A-7	Sequence 1, Appl1
35	757	11.2	1208	9	US-10-213-329-1	Sequence 11, Appl1
36	757	11.2	1218	10	US-09-995-593A-11	Sequence 3, Appl1
37	757	11.2	1218	10	US-09-944-849-3	Sequence 7, Appl1
38	747	11.1	1218	9	US-10-219-248-7	Sequence 7, Appl1
39	747	11.1	1218	9	US-10-219-247-7	Sequence 7, Appl1
40	747	11.1	1218	10	US-09-855-722-7	Sequence 6, Appl1
41	739.5	11.0	566	9	US-09-900-449A-6	Sequence 36, Appl1
42	735.5	10.9	2743	9	US-10-037-182-6	Sequence 2, Appl1
43	735.5	10.9	3695	9	US-10-037-182-2	Sequence 12, Appl1
44	721.5	10.7	310	9	US-10-084-994-12	Sequence 8, Appl1
45	717.5	10.6	1801	10	US-09-938-275-8	

ALIGNMENTS

RESULT 1

US-10-092-390-2
Sequence 2, Application US/10092390
Publication No. US20030013865A1
GENERAL INFORMATION:
APPLICANT: Yu, Xuanchuan
TITLE OR INVENTION: No. US20030013865A1 Human EGF-Family Proteins and Polynucle
FILE REFERENCE: LEX-0317-USA
CURRENT APPLICATION NUMBER: US/10/092,390
PRIOR FILING DATE: 2002-03-06
PRIOR APPLICATION NUMBER: US 60/275,013
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 1140
TYPE: PRT
ORGANISM: homo sapiens
US-10-092-390-2

Query Match	100.0%	Score 6744;	DB 9;	Length 1140;
Best Local Similarity	100.0%	Pred. No. 0;		
Matches 1140;	Conservative	0;	Mismatches	0;
			Indels	0;
			Gaps	0;
QY	1	MVISINSLSTFICLLCHWTGIGTASPLNLEDPVNCCHMSYSTVQESYPPHPDQIYRSC	60	
DB	1	MVISINSLSTFICLLCHWTGIGTASPLNLEDPVNCCHMSYSTVQESYPPHPDQIYRSC	60	
QY	61	TDILWFCTGCTHRRYSRYRAYRGEKTYRRKSQCCPGFYEGSEGMKVPHPCADKCVGRGIA	120	
DB	61	TDILWFCTGCTHRRYSRYRAYRGEKTYRRKSQCCPGFYEGSEGMKVPHPCADKCVGRGIA	120	
QY	121	PNTCCCEGPGWGTNCSSACDDDHMGPHCTSRQCCKNALCNPITGACHCAAGFRWRCE	180	
DB	121	PNTCCCEGPGWGTNCSSACDDDHMGPHCTSRQCCKNALCNPITGACHCAAGFRWRCE	180	
QY	181	RCEGTGANDHOROCNGATCDHVTGECRCPPYTAFCEDLCPPKHGPOCEORPC	240	
DB	181	RCEGTGANDHOROCNGATCDHVTGECRCPPYTAFCEDLCPPKHGPOCEORPC	240	
QY	241	ONGGVCHVTGECSCPSGMMGTVCGPCPEGRFGKNCSEQCHNGTCDATGACGHCSP	300	

```

|||||
Db 241 QNGGVCHHTGECSPSGMMGTVCQPCPEGRGKNCSECCQCHNGTCAAGQCHCSP 300
QY 301 GTGGERCODECPVGTGVLCAETCCQCVNGKCYHVSAGACCEAGFAGERCEATLCEGLY 360
Db 301 GTTGERCODECPVGTGVLCAETCCQCVNGKCYHVSAGACCEAGFAGERCEATLCEGLY 360
QY 361 GIKCDRCPCHELENTSCHPMSEGCACPKWMSGLYCNETCSPGFYEGACQILSCONGAD 420
Db 361 GIKCDRCPCHELENTSCHPMSEGCACPKWMSGLYCNETCSPGFYEGACQILSCONGAD 420
QY 421 CSVTGKCTCAPEFKGIDSTPCPLGTGYNCSRCCKNDVAVSYDSSCTCAKMHGY 480
Db 421 CSVTGKCTCAPEFKGIDSTPCPLGTGYNCSRCCKNDVAVSYDSSCTCAKMHGY 480
QY 481 DCSIRCPSTGWRGCMILTCQCLNGGACNTLDGTCTCAPGRGKCELPDODGTGYNCAE 540
Db 481 DCSIRCPSTGWRGCMILTCQCLNGGACNTLDGTCTCAPGRGKCELPDODGTGYNCAE 540
QY 541 RCDCHADGCHPTTGHCRCLPGMSGVHCDVCAEGRMPCSLPCYCKNGASCSPDDGIC 600
Db 541 RCDCHADGCHPTTGHCRCLPGMSGVHCDVCAEGRMPCSLPCYCKNGASCSPDDGIC 600
QY 601 ECAPGRGTTTCORICSPGFGRHRCSTCPQCVHSSGPRCHHTITGLCCLPEFTALCNEVC 660
Db 601 ECAPGRGTTTCORICSPGFGRHRCSTCPQCVHSSGPRCHHTITGLCCLPEFTALCNEVC 660
QY 661 PSGRFKNAGICTCTNNGCNPIIDRSOCQPYGMISDCSOPCPRAHMGNCIHTCNCHN 720
Db 661 PSGRFKNAGICTCTNNGCNPIIDRSOCQPYGMISDCSOPCPRAHMGNCIHTCNCHN 720
QY 721 GAFCSAYDECKCTPGMTGLYCTORCPPLGFYGDALICQCCONGADCDHISGQCTRTGF 780
Db 721 GAFCSAYDECKCTPGMTGLYCTORCPPLGFYGDALICQCCONGADCDHISGQCTRTGF 780
QY 781 MGRHCKOKPBGTYGGYGCQICDCLNNSTCDHTTCTCYSPGKRGARCOAGIYIGNLN 840
Db 781 MGRHCKOKPBGTYGGYGCQICDCLNNSTCDHTTCTCYSPGKRGARCOAGIYIGNLN 840
QY 841 SLSTRTALPADSYOIGALAGIILVVLFLALFLIYRHKOKGESSMPATYTPAMR 900
Db 841 SLSTRTALPADSYOIGALAGIILVVLFLALFLIYRHKOKGESSMPATYTPAMR 900
QY 901 VYNADYITIGLPHSNGNANSHYFTNPSTHTLTQCATSPHVNRRDMYTVTSKNNQLEY 960
Db 901 VYNADYITIGLPHSNGNANSHYFTNPSTHTLTQCATSPHVNRRDMYTVTSKNNQLEY 960
QY 961 NLKNVAPGRGPGVDCGTGLPADMKHGYLNEIGAFGLDRSYVSKSLKDLGRKSEYNSN 1020
Db 961 NLKNVAPGRGPGVDCGTGLPADMKHGYLNEIGAFGLDRSYVSKSLKDLGRKSEYNSN 1020
QY 1021 CSLSSSENRYATIKDPVILPKSSSECGYEMKSPARRDSPYAEINNSTANRYVEEPT 1080
Db 1021 CSLSSSENRYATIKDPVILPKSSSECGYEMKSPARRDSPYAEINNSTANRYVEEPT 1080
QY 1081 VSVYGVCFNSNGRLSODPDLPRKNSHIPCHYDLPLVRDSSSSPKQEDSGSSSSSSSE 1140
Db 1081 VSVYGVCFNSNGRLSODPDLPRKNSHIPCHYDLPLVRDSSSSPKQEDSGSSSSSSSE 1140

RESULT 2
US-10-092-390-4
; Sequence 4, Application US/10092390
; Publication No. US20030013865A1
; GENERAL INFORMATION:
; APPLICANT: Yu, Xuanchuan
; TITLE OF INVENTION: No. US20030013865A1 Human EGF-Family Proteins and Polynucleotide
; FILE REFERENCE: LEX-0317-USA
; CURRENT APPLICATION NUMBER: US/10/092,390
; CURRENT FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: US 60/275,013
; PRIOR FILING DATE: 2001-03-12

```

```

; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 4
; LENGTH: 586
; TYPE: prt
; ORGANISM: homo sapiens
US-10-092-390-4

Query Match          53.4%; Score 3601; DB 9; Length 586;
Best Local Similarity 100.0%; Pred. No. 1e-168;
Matches 586; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVTLSNCSLFFICLLCHMIGTASPLNLEDPNVCMSHESYSTVQSGYRHPEDQIYITSC 60
Db 1 MVTLSNCSLFFICLLCHMIGTASPLNLEDPNVCMSHESYSTVQSGYRHPEDQIYITSC 60
QY 61 TDLNMFKCTRHRYVSTRYARHGEKMYRRKSCCCGTFESGMCYPRHCAKCVHRCTA 120
Db 61 TDLNMFKCTRHRYVSTRYARHGEKMYRRKSCCCGTFESGMCYPRHCAKCVHRCTA 120
QY 121 PNTCOCEPMGGTNCSSACDGDHMGPHCTSRCCCKNGALCNPTTGACHCAAGFRWRCD 180
Db 121 PNTCOCEPMGGTNCSSACDGDHMGPHCTSRCCCKNGALCNPTTGACHCAAGFRWRCD 180
QY 181 RCEQGTGYNDCHQRCQONGATCDHVTGECRCRPGYTGAFCEBDLCPGRHGPCEQRCPC 240
Db 181 RCEQGTGYNDCHQRCQONGATCDHVTGECRCRPGYTGAFCEBDLCPGRHGPCEQRCPC 240
QY 241 QNGGVCHHTGECSPSGMMGTVCQPCPEGRGKNCSECCQCHNGTCAAGQCHCSP 300
Db 241 QNGGVCHHTGECSPSGMMGTVCQPCPEGRGKNCSECCQCHNGTCAAGQCHCSP 300
QY 301 GTGGERCODECPVGTGVLCAETCCQCVNGKCYHVSAGACCEAGFAGERCEATLCEGLY 360
Db 301 GTGGERCODECPVGTGVLCAETCCQCVNGKCYHVSAGACCEAGFAGERCEATLCEGLY 360
QY 361 GIKCDRCPCHELENTSCHPMSEGCACPKWMSGLYCNETCSPGFYEGACQILSCONGAD 420
Db 361 GIKCDRCPCHELENTSCHPMSEGCACPKWMSGLYCNETCSPGFYEGACQILSCONGAD 420
QY 421 CSVTGKCTCAPEFKGIDSTPCPLGTGYNCSRCCKNDVAVSYDSSCTCAKMHGY 480
Db 421 CSVTGKCTCAPEFKGIDSTPCPLGTGYNCSRCCKNDVAVSYDSSCTCAKMHGY 480
QY 481 DCSIRCPSTGWRGCMILTCQCLNGGACNTLDGTCTCAPGRGKCELPDODGTGYNCAE 540
Db 481 DCSIRCPSTGWRGCMILTCQCLNGGACNTLDGTCTCAPGRGKCELPDODGTGYNCAE 540
QY 541 RCDCHADGCHPTTGHCRCLPGMSGVHCDVCAEGRMPCSLPCYCKNGASCSPDDGIC 600
Db 541 RCDCHADGCHPTTGHCRCLPGMSGVHCDVCAEGRMPCSLPCYCKNGASCSPDDGIC 600

RESULT 3
US-09-796-753-114
; Sequence 114, Application US/09796753
; Publication No. US20030027998A1
; GENERAL INFORMATION:
; APPLICANT: McCarthy, Sean A.
; TITLE OF INVENTION: SECRETED PROTEINS AND USES THEREOF
; FILE REFERENCE: 7853-227-999
; CURRENT APPLICATION NUMBER: US/09/796,753
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 09/183,175
; PRIOR FILING DATE: 1998-10-30
; PRIOR APPLICATION NUMBER: 09/223,094
; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 09/223,546
; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 09/224,246
; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 09/259,388
; PRIOR FILING DATE: 1999-02-26

```

```

; PRIOR APPLICATION NUMBER: 60/122,458
; PRIOR FILING DATE: 1999-03-01
; PRIOR APPLICATION NUMBER: 09/312,359
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 09/336,536
; PRIOR FILING DATE: 1999-06-18
; PRIOR APPLICATION NUMBER: 09/342,687
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: 09/345,464
; PRIOR FILING DATE: 1999-06-30
; PRIOR APPLICATION NUMBER: 09/365,164
; PRIOR FILING DATE: 1999-07-30
; PRIOR APPLICATION NUMBER: 09/399,723
; PRIOR FILING DATE: 1999-09-20
; PRIOR APPLICATION NUMBER: 09/409,634
; PRIOR FILING DATE: 1999-09-30
; PRIOR APPLICATION NUMBER: 09/471,179
; PRIOR FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: 09/474,071
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 09/474,072
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 09/514,010
; PRIOR FILING DATE: 2000-02-25
; PRIOR APPLICATION NUMBER: 09/516,745
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 09/572,002
; PRIOR FILING DATE: 2000-05-14
; PRIOR APPLICATION NUMBER: 09/597,993
; PRIOR FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: 09/599,596
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 09/630,334
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: 09/606,565
; PRIOR FILING DATE: 2000-06-29
; PRIOR APPLICATION NUMBER: 09/606,317
; PRIOR FILING DATE: 2000-06-29
; PRIOR APPLICATION NUMBER: 09/665,666
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: 09/677,751
; PRIOR FILING DATE: 2000-09-30
; NUMBER OF SEQ ID NOS: 162
; SEQ ID NO 114
; LENGTH: 1050
; TYPE: PRF
; ORGANISM: Homo sapiens
US-09-796-753-114

```

```

Query Match      37.2%; Score 2506.5; DB 9; Length 1050;
Best Local Similarity 40.5%; Pred. No. 3.9e-115;
Matches 490; Conservative 111; Mismatches 345; Indels 263; Gaps 30;

```

```

QY 14 LLLCHWIGTASPLNLEDPNVCSHWESYSTVQESYPPHFDQIYYTCTDILN-----66
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 9 LLLAVGRLAGTLNPSDPNCSFEWSESYTTTKESHSPFSLPSEPC--RPMGPHATCP 66
QY 67 -----FKCRHRSYR-----TAY 80
DB 67 SPOTORKLASRDSFCMVCAAGVQWRDRSALOPQTGNALSMRPQPRVLSGAPSLSPGH 126
QY 81 RHGEKTYRRKSQCPPEYSGEKCVPHCADKCVHGRCLAPNTCCPEPGMGNTCSA--138
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 127 TVVVKTDHRRQLOCHGFYSRGFCVPLCAQECVHGRCAVPNOCQCVPGMRGDDCSSAPN 186
QY 139 -----CDGDHMGPHCTSRCKCKNGALCNPLTGACHCAAGFGMRGCEDEGTYGNDCHOR 194
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 187 CLAPCTPGYGPACQFCRCQ--HGAPCDPQTGACFCFPAERTRGPGSDVSCSGT-----237
QY 195 CQCGNATCTCHVTGECRCPPGTGACFEDLCPRGKRGPOCEQRCPCONGGVCHHTVGECS 254
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 238 -----SGFRC-----PSTH-----PCNGGVGTVPQSGSCS 262

```

```

QY 255 CPFGMMGTVCOPPEGRGFRKNCSSQCCCHNGGTCDAATGCHSPGTGTERCODECPVG 314
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 263 CPFGMMGTICSLCPPEGRFHPNCSQECRCNNGGLCDPFTQCCRCARPYTDRCRECPVG 322
QY 315 TYGLCAETCCQVNGKCVHVSAGCLCEAGFAGEBCEARLCPBGLYIKDKRCPCHLEN 374
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 323 RFGQDCAETCDCAAPDANCFPANGACLEHEFTGTDRCTDRLCRDPGFYLSLQAPCTOREH 382
QY 375 THSCHPMGSEACPKPMGSLYCNETCSPGFYGEACQOICSCQNGACDSDYTKCTCAPF 434
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 383 SLSCHPMNGECSCLPGVAGLHCNESCPODTHGECQEHCLCLHGGVCOATSGLOCCAPGY 442
QY 435 KGIDCSPPCLPTGYGNCSSRCGCKNDAYCSPVDGSCSTKAGMHGVNCSIRCPSTGTAFG 494
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 443 TGRHCSLCPDPTTYGVNCSKSCENAIACSPIDGECVCEGMORGNCSVPCPGTWFGS 502
QY 495 CNLTQCLNGACNTLDGTCTCAPWRGECCELPDODGYTLNCAERDCSHADGCHPTT 554
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 503 CNMSQCAHBAVCSPPQGACTCTPRMHAHCQLPCPKGQFGECCASRCDCHSDGCDPVH 562
QY 555 GHRCRLPBMGVCNCDVSCAGRWGPNCSLPCYCKNAGSACSPDDGICCAAGFRGTQRT 614
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 563 GRCQCGAGMAGARCHLSCPGLMGVNCNNTCTCKNGGTCLPENGNCVACAFGRSPQORS 622
QY 615 CSPGFYGRCSQTCPCQCVHSSGCHHTGLCDLPGFTGALCNBVCPSGRFKNACICT 674
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 623 CQFGRYGR-----CVP-----CK 636
QY 675 CTNNGTCNPIDRSQCCYPGMIGSDCSQPCPPAHMPCNICTCMCHNGAFCSAYDECKCT 724
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 637 CANHSFCHPNSNGTCYCLAGWTGPDSCQPCPRGHGEMCACTCCCHHGCTCHPDQSGCTCP 656
QY 735 PCWMTGLCTQRCPLGFGKCALICQONGADCDHIGGCTCRGFGFRCEQKCPSGTY 794
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 697 LGMTGHHLECGCPGTGANCSPCCGPGKEC-----HPE-----732
QY 795 GYGCQRQICDCLNNTCDHNGTCYCSPGMKGARCDQAGVILVGNLSRSTALPAD--852
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 733 -----TGACVCPRHSGAPCR-----IG-----IQEPHYMPTTPV 763
QY 853 SY-QIGAIAGIILVLELLELFIILYRRKOKRESSMPAVTYTTPAMRVVNDYTTISGT 911
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 764 AYSNIGAVIGIAYLSLVALVLELFIYRHMOKKEHHHLVAVYSSG--RLDGSEYVAPDV 822
QY 912 LPHNSGNANSHYTNPSYVTLFOCATSPVNNRDRNTVTKSKNNOFLVULK--VNGCKR 970
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 823 PP-----SYSHYSNPSYHTLSQCSPPPPPK-----VPGPLFASLQNERPG--866
QY 971 GPVG-DCTGTLPADMKH-----GYLNLGAFGLDRSYMGKSL-----KDLGKNSEY 1016
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 867 GAGCHDHTTLPLADMKHRRPPGPIADR--GSSRLDRYSYSYSGPFPYDKGLISREEL 925
QY 1017 NSSNCSLSSENPATIKDPPVILPKSSECGYEMKSPARNDSPYAEINNSTSANRVYE 1076
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 926 GASVAVSL--SEENYATIRDLPSLPGGRESYSYEMKMPGSAAPROPQFWDSDRR--981
QY 1077 VEPYVSYVGVGSNNGSL-----SDP-----DLPKNSHIPPCHYDLDPVRDS 1119
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 982 -QPOPORDSTGYDQPSPLIHDRDSVGSQPLPGLPGRGHYDPSKNSHIPGHYDLPPVRHP 1040
QY 1120 SSSP-KOED 1127
DB 1041 PSPPLRRQD 1049

```

```

RESULT 4
US-09-796-753-100
; Sequence 100, Application US/09796753
; Publication No. US20030027998A1
; GENERAL INFORMATION:
; APPLICANT: McCarthy, Sean A.
; TITLE OF INVENTION: SECRETED PROTEINS AND USES THEREOF
; FILE REFERENCE: 7853-227-999

```

```

1 CURRENT APPLICATION NUMBER: US/09/796,753
2 CURRENT FILING DATE: 2001-03-01
3
4 PRIOR APPLICATION NUMBER: 09/183,175
5 PRIOR FILING DATE: 1998-10-30
6 PRIOR APPLICATION NUMBER: 09/223,094
7 PRIOR FILING DATE: 1998-12-30
8 PRIOR APPLICATION NUMBER: 09/223,546
9 PRIOR FILING DATE: 1998-12-30
10 PRIOR APPLICATION NUMBER: 09/224,246
11 PRIOR FILING DATE: 1998-12-30
12 PRIOR APPLICATION NUMBER: 09/259,368
13 PRIOR FILING DATE: 1999-02-26
14 PRIOR APPLICATION NUMBER: 60/122,458
15 PRIOR FILING DATE: 1999-03-01
16 PRIOR APPLICATION NUMBER: 09/312,359
17 PRIOR FILING DATE: 1999-05-14
18 PRIOR APPLICATION NUMBER: 09/336,536
19 PRIOR FILING DATE: 1999-06-18
20 PRIOR APPLICATION NUMBER: 09/342,687
21 PRIOR FILING DATE: 1999-06-29
22 PRIOR APPLICATION NUMBER: 09/345,464
23 PRIOR FILING DATE: 1999-06-30
24 PRIOR APPLICATION NUMBER: 09/365,164
25 PRIOR FILING DATE: 1999-07-30
26 PRIOR APPLICATION NUMBER: 09/399,723
27 PRIOR FILING DATE: 1999-09-20
28 PRIOR APPLICATION NUMBER: 09/409,634
29 PRIOR FILING DATE: 1999-09-30
30 PRIOR APPLICATION NUMBER: 09/471,179
31 PRIOR FILING DATE: 1999-12-23
32 PRIOR APPLICATION NUMBER: 09/474,071
33 PRIOR FILING DATE: 1999-12-29
34 PRIOR APPLICATION NUMBER: 09/474,072
35 PRIOR FILING DATE: 1999-12-29
36 PRIOR APPLICATION NUMBER: 09/514,010
37 PRIOR FILING DATE: 2000-02-25
38 PRIOR APPLICATION NUMBER: 09/516,745
39 PRIOR FILING DATE: 2000-03-01
40 PRIOR APPLICATION NUMBER: 09/572,002
41 PRIOR FILING DATE: 2000-05-14
42 PRIOR APPLICATION NUMBER: 09/597,993
43 PRIOR FILING DATE: 2000-06-19
44 PRIOR APPLICATION NUMBER: 09/599,596
45 PRIOR FILING DATE: 2000-06-22
46 PRIOR APPLICATION NUMBER: 09/630,334
47 PRIOR FILING DATE: 2000-07-31
48 PRIOR APPLICATION NUMBER: 09/606,565
49 PRIOR FILING DATE: 2000-06-29
50 PRIOR APPLICATION NUMBER: 09/606,317
51 PRIOR FILING DATE: 2000-06-29
52 PRIOR APPLICATION NUMBER: 09/665,666
53 PRIOR FILING DATE: 2000-09-20
54 PRIOR APPLICATION NUMBER: 09/677,751
55 PRIOR FILING DATE: 2000-09-30
56
57 NUMBER OF SEQ ID NOS: 162
58
59 SEQ ID NO 100
60
61 LENGTH: 636
62
63 TYPE: PRF
64
65 ORGANISM: Rattus sp.
66
67 US-09-796-753-100

```

[illegible]

```

0Y 360 PMSGCACKPMSGLYCNERTCSPEBYGACAOOLSCONGADODSDVSTCKYCAPEFKIDC 4399
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 121 PMSGBCSCOPWAGLHNCSECPDTHGAGCCOEHCLJHGGVCLADSLCKRABPYTPHC 1800
0Y 440 STPECLGYGJNCSSRCGCKNDKAVCSPPDGSCTCKAGMHGVDSSIRPSGTWFGMWTG 4999
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 161 ANLCBPMTYGTGJNCSSHCSENAIACSPYDGTCTCKEBGMONGNCSVPCPPGTWFGSNASC 2400
0Y 500 OCLNGACNITDGYCTCAKPMRGEKCELPQODGYTLNLCARCDCSHADDCPTTGHRC 5558
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 241 QCAHEGVCSPTGTACTCTPPGMRGVHCOLPCPKGGFGBGCASVCCDSDSDCDPVHGCRC 3000
0Y 560 LPMGSGVDCDVCACAEGRMPKMSLPCYCKNKGASCSPDGCICECAPGERGTTCORICSPGF 6199
Db 301 QAGMMGTGCHLPCBPBGFMAGNCSNACTCKNGGTCTVPENGNCVCAPRGRGSCQRPCEPGR 3600
0Y 620 YGHRCSOTCPQCVHSSGPHHITGLCDCLPBGFTGALCNENVCPSGRFGKNCAGICTTNG 6799
Db 361 YGKR-----CVP-----CKNNHS 3740
0Y 660 TCNPLDRSCOCYCPMGISDCSOPCPAPMHWGNCIHTCNCHNGAFCSAYDEBCKTPEWTG 7399
Db 375 SCHPBDGTCSCLAGTGTGDCSECPRGHMLKCSOPCCHHGATGCHQDSCVCIPEWTG 4344
0Y 740 LYCTORCPLAGEYGRKCALICOCONGADODHISGOCTCTRGFMGRHCQKCPSGTYGYGR 7999
Db 435 PNCSECCPSRMFGVNCSDLCCDDGEMC-----HPE-----4655
0Y 800 QICDCLNNSSTDHITGTCYCSPPKMGKARGCDOAGYIYGNLNSLRSRTALPADSYOGAI 8559
Db 466 -----TGACVCPGPHSGAKC-----VGSODEFTIMPTS-PVJHNSIGAV 5044
0Y 860 AGIILVLYVELLALFTIYRHKOKGESSMPAVYTPAMRVVNADYTISGTLPHSNGN 9199
Db 505 IGIAYLGLTVLALVALFITYGRHMQKKEHEHLAVAYSIG-RLDSODVYMDVSP-----5579
0Y 920 ANSHFTMPSYHITLQCATSHVANNRDMRYTKSKNNOLVYNLKNVAPGKRGPGDCTGT 9799
Db 558 SYSYHYSNPSYHTLSQCSPPNPPPN-----KIPGQLFVSSQASERPNRHHGRDHAT 6100
0Y 980 LPADWKH 986
    |||:||||
Db 611 LPADWKH 617

RESULT 5
US-09-796-753-124
; Sequence 124, Application US/09796753
; Publication No. US20030027998A1
; GENERAL INFORMATION:
; APPLICANT: McCarthy, Sean A.
; TITLE OF INVENTION: SECRETED PROTEINS AND USES THEREOF
; FILE REFERENCE: 7853-227-999
; CURRENT APPLICATION NUMBER: US/09/796,753
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 09/183,175
; PRIOR FILING DATE: 1998-10-30
; PRIOR APPLICATION NUMBER: 09/223,094
; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 09/223,546
; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 09/224,246
; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 09/259,388
; PRIOR FILING DATE: 1999-02-26
; PRIOR APPLICATION NUMBER: 60/122,458
; PRIOR FILING DATE: 1999-03-01
; PRIOR APPLICATION NUMBER: 09/312,359
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 09/336,536
; PRIOR FILING DATE: 1999-06-18
; PRIOR APPLICATION NUMBER: 09/342,687
; PRIOR FILING DATE: 1999-06-29

```

US-09-796-753-124

Sequence 124, Application US/09796753

Publication No. US20030027998A1

GENERAL INFORMATION:

APPLICANT: McCarthy, Sean A.

TITLE OR INVENTION: SECRETED PROTEINS AND USES THEREOF

FILE REFERENCE: 7853-227-999

CURRENT APPLICATION NUMBER: US/09/796,753

CURRENT FILING DATE: 2001-03-01

PRIOR APPLICATION NUMBER: 09/183,115

PRIOR FILING DATE: 1998-10-30

PRIOR APPLICATION NUMBER: 09/223,094

PRIOR FILING DATE: 1998-12-30

PRIOR APPLICATION NUMBER: 09/223,546

PRIOR FILING DATE: 1998-12-30

PRIOR APPLICATION NUMBER: 09/224,246

PRIOR FILING DATE: 1998-12-30

PRIOR APPLICATION NUMBER: 09/259,388

PRIOR FILING DATE: 1999-02-26

PRIOR APPLICATION NUMBER: 60/112,458

PRIOR FILING DATE: 1999-03-01

PRIOR APPLICATION NUMBER: 09/312,359

PRIOR FILING DATE: 1999-05-14

PRIOR APPLICATION NUMBER: 09/336,536

PRIOR FILING DATE: 1999-06-18

PRIOR APPLICATION NUMBER: 09/342,667

PRIOR FILING DATE: 1999-06-29


```

1 PRIOR APPLICATION NUMBER: 09/345,464
2 PRIOR FILING DATE: 1999-06-30
3 PRIOR APPLICATION NUMBER: 09/365,164
4 PRIOR FILING DATE: 1999-07-30
5 PRIOR APPLICATION NUMBER: 09/399,723
6 PRIOR FILING DATE: 1999-09-20
7 PRIOR APPLICATION NUMBER: 09/409,634
8 PRIOR FILING DATE: 1999-09-30
9 PRIOR APPLICATION NUMBER: 09/471,179
10 PRIOR FILING DATE: 1999-12-23
11 PRIOR APPLICATION NUMBER: 09/474,071
12 PRIOR FILING DATE: 1999-12-29
13 PRIOR APPLICATION NUMBER: 09/474,072
14 PRIOR FILING DATE: 1999-12-29
15 PRIOR APPLICATION NUMBER: 09/514,010
16 PRIOR FILING DATE: 2000-02-25
17 PRIOR APPLICATION NUMBER: 09/516,745
18 PRIOR FILING DATE: 2000-03-01
19 PRIOR APPLICATION NUMBER: 09/572,002
20 PRIOR FILING DATE: 2000-05-14
21 PRIOR APPLICATION NUMBER: 09/597,993
22 PRIOR FILING DATE: 2000-06-19
23 PRIOR APPLICATION NUMBER: 09/599,596
24 PRIOR FILING DATE: 2000-06-22
25 PRIOR APPLICATION NUMBER: 09/630,334
26 PRIOR FILING DATE: 2000-07-31
27 PRIOR APPLICATION NUMBER: 09/606,565
28 PRIOR FILING DATE: 2000-06-29
29 PRIOR APPLICATION NUMBER: 09/606,317
30 PRIOR FILING DATE: 2000-06-29
31 PRIOR APPLICATION NUMBER: 09/655,666
32 PRIOR FILING DATE: 2000-09-20
33 PRIOR APPLICATION NUMBER: 09/677,751
34 PRIOR FILING DATE: 2000-09-30
35 NUMBER OF SEQ ID NOS: 162
36 SEQ ID NO 124
37 LENGTH: 636
38 TYPE: PRT
39 ORGANISM: Raultus sp.
40 JS-09-796-753-124

```

```

QY 518 PGMREKELPCQDGTGLNCAERDCSHADGCHPTTGCRCLPMSGVHCDVCAEGRW 577
DB 300 LGWTRHECLACPPERYAACHLESCHNNSTCEPATGTGCRGPPFYQACEHPCPGPH 359
QY 578 GPNCLSPCYCKNAGASCSPDDGICECAPGFRGTTCORICSPGYGRCSOTCPQCVHSSGP 637
DB 360 GAGCGGLCMQCGACGACDIPISGRCLCPAFHGHFCRCEPESFGSGCHQRC-DC-DGAP 417
QY 638 CHHTGLDCLDLPFTGALCN 657
DB 418 CDPVTGLCLCPGRSGATCN 437

RESULT 7
US-09-944-849-2
; Sequence 2, Application US/09944849
; Patent No. US20020151487A1
; GENERAL INFORMATION:
; APPLICANT: Nickoloff, Brian
; APPLICANT: Miele, Lucio
; TITLE OF INVENTION: METHOD AND REAGENTS FOR EPITHELIAL BARRIER FORMATION AND TREATMEN
; FILE REFERENCE: 212583
; CURRENT APPLICATION NUMBER: US/09/944,849
; PRIOR FILING DATE: 2001-08-31
; PRIOR APPLICATION NUMBER: US 60/229,614
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patent version 3.1
; SEQ ID NO 2
; LENGTH: 2444
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: misc-feature
; LOCATION: (891)..(891)
; OTHER INFORMATION: The 'Xaa' at location 891 stands for Gly, or Ala.
; NAME/KEY: misc-feature
; LOCATION: (1763)..(1763)
; OTHER INFORMATION: The 'Xaa' at location 1763 stands for Gln, Arg, Pro, or Leu.
; LOCATION: (1787)..(1787)
; OTHER INFORMATION: The 'Xaa' at location 1787 stands for Thr, Ala, Pro, or Ser.
; OTHER INFORMATION: Constitutively Active No. US20020151487A1ch-1
US-09-944-849-2

Query Match 15.38; Score 1034.5; DB 10; Length 2444;
Best Local Similarity 25.88; Pred. No. 4.6e-43;
Matches 316; Conservative 83; Mismatches 304; Indels 523; Gaps 73;

QY 94 CCGFYESGECVPHCADKCVHRC-----IAPNTCCQCPKNGTGCSSACDGDH 143
DB 89 CALGF--SGPLCLPPLNACLTPNCRNGTCDLLTLEYKCRCPGWSGKSCQA----- 141
QY 144 WGRHCTSRCCCKNGALCNPTGA--CHCAAGFRGMRCE--DRCEQG---TYGNDCHQ- 193
DB 142 --DPCASN-PCANGGCLPFEASVICHGPPSFHPTCRQDVNDEGQAPRLCRHGTCNE 198
QY 194 ----RC-----OCONGATC---DHVTEGRCRPPGYTGAFC- 222
DB 199 VGSYRCVCRATHNGPNCERPRVPCSPQNGGTCRPTGVTGTHCACLPGFTGNCSENI 258
QY 223 DLCPG--KHGPOC-----DQRCP-----CONGVCHHTG- 251
DB 259 DDCPGNNCKNGACVGVNTYNNCCPPEMTGQCTEDVDECOQMPNACQNGGTCNHTHG 318
QY 252 -ECSGSPGNGTVCGQ-----PCPEGFRGKNC--SQDC-- 281
DB 319 YNCVGVNGWGEPCSENIIDOCASACRHNATCHDRVASFCCEPHGHTGLCHLNDACIS 378
QY 282 -QCHNGSTCA--ATGO--CHCSFGYGERGQ--DECPRVGYGVLCALCTCCQVNGSKCY 333
DB 379 NPNEGSNCQDTPNPNVNGKAICTCPSSGYTPACSDVDDECSLGAN-----PCBHAKCI 430

```

```

QY 334 HVSGA--CLCEAGFNGERCEARLCEPLGYLKCQRK---PCHLEHTSCHPMSE--CA 386
DB 431 NTLGSEFECQCLQGYTPRCELDV-----NECSNRC--QNDATCLDQIEFGCM 477
QY 387 CRPGHSLYC-----NE---TCSPGFYGEACQ---QICS---C 415
DB 478 CMPEYGVHCEVNTDECASSPLCHNGRCLDKRINEFQCEBPGFTGHLCQYDVDECASTPC 537
QY 416 ONGADC-----DSV--TGKCTCAPGPKG 436
DB 538 KNGAKCLDGPMTYTCVTEGTGTHCEVDIDECDDPDCYSCCKDGVATFTCLCPGYTG 597
QY 437 IDGSP-----PCPL-----GTGINS-----SRGCKNDAYCS 465
DB 598 HHCETININESSPQCRNLNGTQDPPDMAVLCFLAGTTPGNEINDDOCASPCDSGTCLD 657
QY 466 PVDG-SECTCRKAGHWGVDCSIR-----CPSGTWFGQNL-----TC----- 499
DB 658 KIDYECACEPQYTGSMCNSNIDECAGNPNCHNGGTCEBDGINGFTCRGPEGYHDEPTCLSEV 717
QY 500 -----OCLNGACNTLDG--TCTCAPGMRGKCEL----- 527
DB 718 NECSNRCVHACRDSLNGYKCDPDPNGSNGTNCINNNECSNPNVNGTCKMTSGIYC 777
QY 528 PCQDGTGLNCAERCD-----CSHADG--HTTGH--CRCLPGSVHCDV---CAEG- 575
DB 778 TCRGFSGPNQOTNINECASNPCLNKGTICIDVAGYKNCNCLPYTGATCEVYLAFCAPSP 837
QY 576 -RMGPNC-----SLPYC---KNGASCSPDDGICECAPGFRGTTCQRT-----CS 616
DB 838 CRNGEGRQSDYSEFSVCVCTPAGAKGOTCEVDINECVLSCRHGASCQNTGHYRCHQ 897
QY 617 PGFYGHRSQTCPCQVHSSGPPCH-----ITGLDCLPFTGALCNE----- 658
DB 898 AGYGRNCEFTDIDC--RPNCHNGSGCTGDSINTAFDCPLGFRGTCEEDINECASPDC 955
QY 659 -----VPSGFRGKNCAG--TCF---CTNNGTCNPIDR---SCQCYG 693
DB 956 RGNAGCTDCVDSYTCCTPAGSGIHCEMNPDCYESSQFNQGT--VDGINSFYCLCPG 1013
QY 694 WIGSDC-----SOP-----CPAHNGPNC--IHNCN---CHNGA 722
DB 1014 FTGSGCHVYNECDSRCLLGGTCQDGRGLHRTCTPGGTGPNCONLVMHCDSSPKNGG 1073
QY 723 FC-----SAYDECKCTPGMTGLYCTO----- 744
DB 1074 KCMQTHQY--RCBCPSGWTGLYCDVPSVCEVAHQGVYARLQHGGLCVDAAGTTH 1131
QY 745 -RCPLEFGYKDALI---CO-----CONGADC-DHISG--QCTCRGFMGRHCEQK----- 788
DB 1132 CRCAGYTSYCEDLVDEGSPSPCONGATCTDYLGYSCKAVAGHYVNCSEIDECLSH 1191
QY 789 -----CPSGTGYGCRQICD-----CLNNSNC-DHITG 815
DB 1192 PCQNGGTCLDLPNTYKSCSPRGTOYHCEINVDCCNPVDPVSRKPCFNNGTCVDVGG 1251
QY 816 -TCYCSPGMKGARCDAQVIIVGNTLN 840
DB 1252 YSCTCPGPFVGERCE-----GDVN 1270

RESULT 8
US-10-084-994-8
; Sequence 8, Application US/10084994
; Publication No. US20030023070A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: Attraction-like Polynucleotides, Polypeptides, and Antibodies
; FILE REFERENCE: PT011P1C1
; CURRENT APPLICATION NUMBER: US/10/084,994
; PRIOR FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/790,621

```

```

? PRIOR FILING DATE: 1999-12-29
? PRIOR APPLICATION NUMBER: 09/474,072
? PRIOR FILING DATE: 1999-12-29
? PRIOR APPLICATION NUMBER: 09/514,010
? PRIOR FILING DATE: 2000-02-25
? PRIOR APPLICATION NUMBER: 09/516,745
? PRIOR FILING DATE: 2000-03-01
? PRIOR APPLICATION NUMBER: 09/572,002
? PRIOR FILING DATE: 2000-05-14
? PRIOR APPLICATION NUMBER: 09/597,993
? PRIOR FILING DATE: 2000-06-19
? PRIOR APPLICATION NUMBER: 09/599,596
? PRIOR FILING DATE: 2000-06-22
? PRIOR APPLICATION NUMBER: 09/630,334
? PRIOR FILING DATE: 2000-07-31
? PRIOR APPLICATION NUMBER: 09/606,565
? PRIOR FILING DATE: 2000-06-29
? PRIOR APPLICATION NUMBER: 09/606,317
? PRIOR FILING DATE: 2000-06-29
? PRIOR APPLICATION NUMBER: 09/665,666
? PRIOR FILING DATE: 2000-09-20
? PRIOR APPLICATION NUMBER: 09/677,751
? PRIOR FILING DATE: 2000-09-30
? NUMBER OF SEQ ID NOS: 162
? SEQ ID NO 116
? LENGTH: 497
? TYPE: PRT
? ORGANISM: Mus musculus
? OS-09-796-753-116

```

[illegible]

```
QY 1106 HIPCHYDLLPVRDSSSSPKQ 1125
      ||| ||| ||| | | :
Db 474 HIPGHYDLLPVRHPPSPPSR 493
```

RESULT 10
HC-00-070 750 134

```

: Sequence 134, Application US/09870759
: Patent No. US20020177551A1
:
: GENERAL INFORMATION:
:
: APPLICANT: TERMAN, David S
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT OF NEOPLASTIC DISEASE
: FILE REFERENCE: 870759
: CURRENT APPLICATION NUMBER: US/09/870,759
: CURRENT FILING DATE: 2002-01-14
: PRIORITY APPLICATION NUMBER: US 60/208,128
: PRIORITY FILING DATE: 2000-05-30
: NUMBER OF SEQ ID NOS: 166
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 134
: LENGTH: 830
: TYPE: PRT
:
: ORGANISM: Homo sapiens
: US-09-870-759-134

```

Query Match	12.9%	Score 870.5;	DB 9;	Length 830;
Best Local Similarity	28.4%;	Pred. No. 1.9e-35;		
Matches 183; Conservative	45;	Mismatches 154;	Indels 263;	Gaps 22

```

0Y      93  OCCPPEYEGEMC -VPHC - ADKCVHG - CIAPIPTCOCEPBMGTGNCSSACDDHMRPHC 148
0Y      40  OCCAGHRQADDDCTTPICEBGRDACKDEYCAKPGLCRCKPGEFFAHNCSNRPGQWYBDC 99
0Y      149  T5RCCKNGALCNPIITGACACAAEGRMGRCEBDRCEOGTYGNDCHQRCOQNGATGACDVTG 208
Db      100  RESCECHPHGOCERPATGACCOADRMGARGCEPCACSPHGR-----CDPATG 146
0Y      209  EGRCPRTGTAFCEBDCLEPRKHHGRQCEBRCFQNGVGNHHTGBCSCPSCMGTVGQPC 268
Db      147  VCHCE-----GWSSTCRRP- 162
0Y      269  PEGRGKNCOSQECCHNGGT -CDATGQCCHSPGYTGRCODECPVGTGYVLAEITQCV 327
Db      163  -----CQCNTHAAACBEGATGACVCKPBMGR- 189
0Y      328  NGKCTYHVSAGCLCEAGFAGERCEARCLPESLYGIKCDKRCPLHENTHSCHPMSEGCAC 387
Db      190  -----CSFRONCH- GSPCEDDSGRCAC 210
0Y      388  KPGMSGLYCNETCSPGFGEACOOIGSONADDSVTGKCTCAPRKIGDSTPRLGT 447
Db      211  RRGW-----WGRPCQOQCELVKRR -CSAASGETCPGPGRGARCELPSPAGS 256
0Y      448  YGINCSSRCG -CKNDAYCSPVDSQCTCKAGHGVDSICRPSGTWFGGCMILTCQLNGGA 506
Db      257  HGVOCAHSCGCRKHNPECSPTGSC- 282
0Y      507  CNTLDGTCTCAPGRMGEKCELPQDDGTGYGLNCAER -DCSHADSCAPTTHGC -RCLPGRMS 564
Db      283  -----SCBEPGNNGTCCQDPCPLRGTFFGSECCEDQCHCHHNGEACEPDTTHCOCROPBGW- 333
0Y      565  GVHCDVYCAEBRMQPNCSILPYCKKNGASCSPPDDICBACARPGTTCQRI CSPGFYGHNC 624
Db      334  -----LQPRCEDP-----CPTGTGEBDC 351
0Y      625  SOTPCPYHSSGRCNHTTGLDCLPGETGALCNEVCPSGRFGKNCAGICTNNGTGNCPI 684
Db      352  GSTCPTCV -OGSCDFTVTDGVCVCSAGYWGSPSCNAPSAGFHHNNCSVPBCB -EGLCHPY 408
0Y      685  DRSCQCYPG-----WIGSDCSQPCPPAHNRP 710
Db      409  SGSCD--PGSGSRPTALVGSVLPLLLTFLGLAC--CACCCNAP 448

```

RESULT 11
US-09-870-759-140
; Sequence 140, Application US/09870755
; Patent No. US20020177551A1

Query Match	12.9%	Score 870.5;	DB 9;	Length 830;
Best Local Similarity	28.4%;	Pred. No. 1.9e-35;		
Matches 183;	Conservative 45;	Mismatches 154;	Indels 263;	Gaps 22;

QY	93	OCCEFEYSGEHC-VPHC--ADKCVHGR-CLAPNTCCCEPMTGGTNCSSACDDDHMGPHC	148
Db	40	QCCAGHROKDEBCTIPTEGRDACCOKDEVCYKPKLCRKPFEFFAHSSSRCPQYGM-PDC	99
QY	149	TSRCCCKNGALCNPTTGACHCACAGRGWRCEBDRCEBGTGYNDCIHQRCCONGATCQHVITG	208
Db	100	RESCFCHPHGOCCEPATGACACCQADBRMGARCEFFPCACBPGR-	146
QY	209	BERCEPRTGTGAFCEDLCPRKHHGQCCEBRCCQNGVCHHYTERSCSPSGMGTVCQPC	268
Db	147	VCHCEP-	162
QY	269	PEGRGKNCSEOCCHNGGT-CDATGQCCHCSPGYTERCODECPVGTGYVLCAETCCQCV	327
Db	163	CCQNTAAACEDATGATGACVCKPEWMKR-	189
QY	328	NGCKTYHVSACILCBAGRGERCEARLCPBGLYGIKCDKRKPCQHLNTHSCHPMSEAC	387
Db	190	CSFRCNCH--GSPCEBDSGRCAC	210
QY	388	KPGMSGLYCNETSPGFGEACQOICSQONGADDSTVGCTCQAPRGKIDSTPCPLGT	447
Db	211	WRPCCQOCECVKRR-CSAASGECTCPRGGRGARCLPCPAPS	256
QY	448	YGINCSSRCG-CKNDAYCSPVDGSCCTKAGWGVDCISRPCSTGWFGCALTCQCLNGA	506
Db	257	HGVQCAHSCGCRKHNEPCSPDTGSC-	282
QY	507	CNTLDGTCAPRGRGEXCELPQODGTGYLGNLCNER-DCSHADGHPRTGHC-RCLPGNS	564
Db	283	SCBPNNNGTCCQOCPCLPRGTGSGSCBQCCPHCNHGAACEPBDTHCCRCRDPGW-	333
QY	565	GVHCDVCAERGMPCNSLPCYCKNGASCSRPBDGICBACAGFRGTCQORICSPGFVGHNC	624
Db	334	LQPRCBP-	351
QY	625	SGCPQCVHSSGPCNHTITGLDCLPRFTGALCNEVSPSGFRGKNCAGICTTNGTNCNP	684
Db	352	GSCTPCTCV--QGSCTDVTGDCVCSAGYWGPPSCNASCAPAGHNGNCSVPBCEP-EGLCPIV	408
QY	685	DRSCQCTPG--WIGSCQSQCPRAHHGP	710
Db	409	SGSCQ--PGSGSRDTALIASLVPLLLFTGLAC--CACCCWAP	448

RESULT 12
US-09-842-758-57

Sequence 57 Application US/09842758
Publication No. US2003008324A1

GENERAL INFORMATION:

APPLICANT: Vernet, Corine A. M.
APPLICANT: Fernandes, Elma R.
APPLICANT: Gerlach, Valerie
APPLICANT: Shinkets, Richard A
APPLICANT: Malvankar, Uriel M
APPLICANT: Boldog, Ferenc L
APPLICANT: Zernhusen, Bryan D
APPLICANT: Spytek, Kimberly A
APPLICANT: Majumder, Kundu
APPLICANT: Tchernev, Velizar T
APPLICANT: Padigarau, Murallidhara
APPLICANT: Patturajan, Meera
APPLICANT: Burgess, Catherine E
APPLICANT: Gangoli, Esna A
APPLICANT: Smithson, Glenda
APPLICANT: Rastelli, Luca
APPLICANT: MacDougall, John R
APPLICANT: Taupier, Raymond J
APPLICANT: Grose, William M
APPLICANT: Edward, Szekeres S
APPLICANT: Alsbrook II, John P

TITLE OF INVENTION: No. US2003008324A1 Proteins and Nucleic Acids Encoding Same

FILE REFERENCE: 1596-783

CURRENT APPLICATION NUMBER: US/09/842,758

CURRENT FILING DATE: 2001-04-25

PRIOR APPLICATION NUMBER: 60/200,158

PRIOR FILING DATE: 2000-04-26

PRIOR APPLICATION NUMBER: 60/200,613

PRIOR FILING DATE: 2000-04-28

PRIOR APPLICATION NUMBER: 60/200,780

PRIOR FILING DATE: 2000-04-28

PRIOR APPLICATION NUMBER: 60/201,006

PRIOR FILING DATE: 2000-05-01

PRIOR APPLICATION NUMBER: 60/201,007

PRIOR FILING DATE: 2000-05-01

PRIOR APPLICATION NUMBER: 60/201,236

PRIOR FILING DATE: 2000-05-01

PRIOR APPLICATION NUMBER: 60/201,238

PRIOR FILING DATE: 2000-05-01

PRIOR APPLICATION NUMBER: 60/201,186

PRIOR FILING DATE: 2000-05-02

PRIOR APPLICATION NUMBER: 60/201,474

PRIOR FILING DATE: 2000-05-03

PRIOR APPLICATION NUMBER: 60/201,508

PRIOR FILING DATE: 2000-05-03

PRIOR APPLICATION NUMBER: 60/220,591

PRIOR FILING DATE: 2000-07-25

PRIOR APPLICATION NUMBER: 60/232,678

PRIOR FILING DATE: 2000-09-15

PRIOR APPLICATION NUMBER: 60/263,217

PRIOR FILING DATE: 2001-01-22

PRIOR APPLICATION NUMBER: 60/265,160

PRIOR FILING DATE: 2001-01-30

NUMBER OF SEQ ID NOS: 113

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 57

LENGTH: 830

TYPE: PRT

ORGANISM: Homo sapiens

US-09-842-758-57

Query Match 12.9%; Score 870.5; DB: 9; Length 830;
Best Local Similarity 28.4%; Pred. No. 1.9e-35;
Matches 183; Conservative 45; Mismatches 154; Indels 263; Gaps 22;

QY QCCGGFYSAGMC-VPHC--ADKCVHGR-CIAPTNCOCPEPGMGGNCSSACDGHMPPHC 148
||| : : || - | : | : || : | : || : | : || :
Db QCCAGWRKDKDECHTIPICEGDACQKDEVCVCFGLCKCRKPFFGAHGCSRCRPGQYWEPDC 99

149 TSRCQCKNGALCNPIPTGACHCAAGFGWRCEDRCEQGTYGNDCHQRCCQNGATCDHYTG 208

```

Db      100 RESCHPHHGQCEPATIGACQQCADRMGAKERPPCACGRHR-----COPATG 146
QY      209 ECRCPFGYTGAFCEDLCPGKKGHPQCEQRPCQNGGVGHVHTGECSGPSGMGTVCGPC 268
Db      147 VCHCEP-----              ||| :|||
QY      269 PEGRFRKNCSDEQCCHNGST -CDATGQCHSPGYTGERQODEPVGTGYVLCAETTCQV 327
Db      163 -----CQCNTAARACEOATGACVCKRGWMMGR-----CWMSSTCRRP- 162
QY      328 NGKRKYHVSAGACLCEAFGRAGERCEARLCPEGLYGICDKRPCHELNTHSCHPMSEGCAC 387
Db      190 -----CSFRCNCH-----GSPCEDSGRCAC 210
QY      388 KPGMGIYCNETCSGFYEACQQLSCONGADDSVTGKCTCAPRFKJIDGSTPCPLGT 447
Db      211 RFGW-----WPRECOQOCCECYRGR-CSAAISECTCPPRFGRKARCLPCBPAGS 256
QY      448 YGINSSRCQG-CRNDAYVSPVDGSCTCKAGMHGVDCSIROPSGTWGFGCNLTCQCLINGA 506
Db      257 HGVOAHASGCRKKHNPRCSLPDTGSCB-----                282
QY      507 CNTLDGTCTCAGMGKEKCELPCQDGTGYLNCABERC-DCSHADGCHPTTGHC-RCLPWGS 564
Db      283 -----SCEFGMWGTQOQDPCLPDTFGESEQDQPCRKHNRGACBEDTHCCRCDBGW- 333
QY      565 GVHCDSVCAEGRMNCSLRPCYCKKNAGASCSPDDGICECAPRGFTGTQQRICSPGFYGNRC 624
Db      334 -----LGFRCEBP-----CPIGTGFEDEC 351
QY      625 SOTPCOVHSSGPCHHTITGLDCLDPFTGALCNENWCPSGRFRGNKAGICTNTNGTCNP1 684
Db      352 GSTCPTCV--QGSODTVTGDDCYCSAGIWMGVRPCSNASCAPAFHGNKNVSPCCBP-EGICHVP 408
QY      685 DRSCCCTPG-----WTGSDCSQPCPPAHNGP 710
Db      409 SGSCQ--PEGSGSRDTALIYGLVPLLFLTLGLAC---CACCCWAP 448

RESULT 13
US-09-900-449A-4
; Sequence 4, Application US/09900449A
; Publication No. US20030040616a1
; GENERAL INFORMATION:
; APPLICANT: ZHONG, Jenny et al.
; TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND
; FILE REFERENCE: CLO01271
; CURRENT APPLICATION NUMBER: US/09/900,449A
; CURRENT FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 639
; TYPE: PRP
; ORGANISM: Strongylocentrotus purpuratus
US-09-900-449A-4

Query Match          12.1%; Score 815; DB 9; Length 639;
Best Local Similarity 30.9%; Pred. No. 7, 8e-33;
Matches 236; Conservative 69; Mismatches 242; Indels 216; Gaps 58;

QY      154 CKNALCLNPITYA----CHCAAGFRKWRC--DRDQGTGYGDCHQRCCQNGCATC-DH 205
Db      5 CLNGVIC--VVGVMNFECTCLAGFTGVRCENVINDECAS-----APCQNGGICIDG 52
QY      206 VTG-ECRCPFGYTGAFCEDLCPGKKGHPQCEQRPCQNGGVGHVHTGE--CSCPSPGMGT 262
Db      53 INGTYCSEPFLGSGDNCE-----NDDECSS-IPLNGTGYDVLYNATMYCVCAPRWTPGR 105
QY      263 VCQGPCPEGRFRKNCSQECQHNGGTC-DAATG-QCHCSPGYTGERCQ---DECPVGTGY 317

```

```

Db 106 TCADNIDE-----CA-SAPCGGVGICIDVNGVYMDCCPGYGTGTCETDIEC----- 152
Qy 318 VLCAETCQCGNGKCYH-VSG-ACLEAGFAGRCFARLCPEGLYGIKDKRCPCHLENT 375
Db 153 ----APPQNGSGDCVDVNGVYICAPFEDLNCE-----NNI 187
Qy 376 HSC--HPMSGECAKPGMSGLYCNETCSPEFYGEACQ---QICS---CONGADC-DSVTG 426
Db 188 DECAHRPCONGAVCVGVNGVYVC--TCSAGYGVLCETDINECASMPCLNGGVCTDLVNG 245
Qy 427 K-CTCAKPKGIDCSFPCPLGTGNGVYVCAADPVSSTCEFDREDECAHAPCLNGACADVYN 482
Db 246 YICTCAGFEGINCEITD-----BCAS-FPCQNGATCTDVNGVYCVPGYGTGLCE 298
Qy 483 -----SIRCPG-----TWFGCNL-----TCQ-----CLNGGAC-NTLD 511
Db 299 TDINECASFPCLNGGTCNDQVNGVYVCAADPVSSTCEFDREDECAHAPCLNGACADVYN 358
Qy 512 G-TCTCAPGMRGEKCEL-----PCQDQTYGLNCAERCDSSHADGCHPTTGHCRLPG 562
Db 359 GFYCTCLPMEGENCEINTEDECASSPCMNG--GL-CVDQVN-SYV-----CFCLPG 405
Qy 563 WSGVHDSVCAEGRMGNPSLPCYCKNGASCSPDDGICECAPGFRGTTCQRTICSPFYGH 622
Db 406 FTGHCGETIDECASSP-----CLNGGQCIDRDVSYEC-----VCAAGYAV 447
Qy 623 RSQSTPCQCVHSSGPCNH-----ITG-LCDLPGFTGALC-NEVCPGSRFGKNCAGIC 673
Db 448 RQGINDEC--ASAPQNGGVYCVGVNGVYVCAADPVSSTCEFDREDECAHAPCLNGACADVYN 497
Qy 674 TCTNNGTCTPI--DRSCQCTPGHIGSDSCOPCPRAHMGPNCHTCHNGHAGCS-AYDGE 730
Db 498 PCLNGGACIEVNGVYVCAADPVSSTCEFDREDECAHAPCLNGACADVYN 550
Qy 731 -CKCPGWTGIVCTQRCPLGFGKDCALICQCONGADC-DHISG-QCTRTGFMGRHCQ 787
Db 551 ICACVGFEGSGNCEINTE-----DECA-SDPCLNGGICVDVNGVYVCAADPVSSTCEFDREDECAHAPCLNGACADVYN 603
Qy 788 KCPGSGTYGVCQRCQICDCLNNSCTDHTG--TCYCSPGMKGAR 828
Db 604 SLDA-----CRSM-PCQNGATCVNNGVYVCAADPVSSTCEFDREDECAHAPCLNGACADVYN 639

RESULT 14
US-09-842-758-20
; Sequence 20, Application US/09842758
; Publication No. US20030083244A1
GENERAL INFORMATION:
APPLICANT: Vernet, Corine A. M.
APPLICANT: Fernandes, Elma R.
APPLICANT: Gerlach, Valerie
APPLICANT: Shinketa, Richard A
APPLICANT: Malvankar, Uriel M
APPLICANT: Boldog, Ferenc L
APPLICANT: Zernusen, Bryan D
APPLICANT: Spytek, Kimberly A
APPLICANT: Majumder, Kumud
APPLICANT: Tchernev, Velizar T
APPLICANT: Padigaru, Muralidhara
APPLICANT: Patturajan, Meera
APPLICANT: Burgess, Catherine E
APPLICANT: Gangolli, Esha A
APPLICANT: Smithson, Glenda
APPLICANT: Rastelli, Luca
APPLICANT: MacDougall, John R
APPLICANT: Taupier, Raymond J
APPLICANT: Grosse, William M
APPLICANT: Edward, Szekeres S
APPLICANT: Alsobrook II, John P
TITLE OF INVENTION: No. US20030083244A1 Proteins and Nucleic Acids Encoding Same
FILE REFERENCE: 15966-783
CURRENT APPLICATION NUMBER: US/09/842,758

```

```

CURRENT FILING DATE: 2001-04-25
PRIOR APPLICATION NUMBER: 60/200,158
PRIOR FILING DATE: 2000-04-26
PRIOR APPLICATION NUMBER: 60/200,613
PRIOR FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: 60/200,780
PRIOR FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: 60/201,006
PRIOR FILING DATE: 2000-05-01
PRIOR APPLICATION NUMBER: 60/201,007
PRIOR FILING DATE: 2000-05-01
PRIOR APPLICATION NUMBER: 60/201,236
PRIOR FILING DATE: 2000-05-01
PRIOR APPLICATION NUMBER: 60/201,238
PRIOR FILING DATE: 2000-05-01
PRIOR APPLICATION NUMBER: 60/201,186
PRIOR FILING DATE: 2000-05-02
PRIOR APPLICATION NUMBER: 60/201,474
PRIOR FILING DATE: 2000-05-03
PRIOR APPLICATION NUMBER: 60/201,508
PRIOR FILING DATE: 2000-05-03
PRIOR APPLICATION NUMBER: 60/220,591
PRIOR FILING DATE: 2000-07-25
PRIOR APPLICATION NUMBER: 60/232,678
PRIOR FILING DATE: 2000-09-15
PRIOR APPLICATION NUMBER: 60/263,217
PRIOR FILING DATE: 2001-01-22
PRIOR APPLICATION NUMBER: 60/265,160
PRIOR FILING DATE: 2001-01-30
NUMBER OF SEQ ID NOS: 113
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 20
LENGTH: 865
TYPE: PRT
ORGANISM: Homo sapiens
US-09-842-758-20

Query Match 12.0%; Score 808; DB 9; Length 865;
Best Local Similarity 33.1%; Pred. No. 2.2e-32;
Matches 156; Conservative 42; Mismatches 164; Indels 110; Gaps 17;

Qy 94 CCRPGYSESEMC-VHCA--DKCVHGR-CIAANTQCCFEGMGKTSSASACDDHNGPHCT 149
Db 64 CCAGWRQOQDEGCIAGVEGNSSTCSSENYCVRGECRCRGRYGANCDDTKRPGFPGPCK 123
Qy 150 SRCQCKNGALCNPTIGACHCAAGFRGWRCEDEGQTYGNDCHORQCONGATCDHVTGE 209
Db 124 ELCSCHPHQGCEDVYGGCTCHA--RRW-----GARCEHACQCOHG-TCHPRSGA 169
Qy 210 CCRPGYTGAFCEDELCPPEKHPQCEQRCPONGGVCHHVTGECSPSGMGTVGQPCP 269
Db 170 CCEPGWMA-----QASACYCSATSRCDPOTGACLHAGW----- 206
Qy 270 EBRPFKNGSQEGCHNGTCDAAATGCHQCHSPYTGEBRODECPVATGYLCAETGCVNG 329
Db 207 ---WGRSCNNQAC--NSSPCEBOGSRQCR-----ER-----TFGARCRDYCCRG 249
Qy 330 GRCYHVSACLEAGFAGRCFARLCPEGLYGIKDKRCPCHLENTHSCHPMSGECAK 389
Db 250 -----RCHPYDGCACBP 262
Qy 390 GWSGLYCNETGSPFYGEACQOIC--SCONGADCDSVTGKC-TCAPGFGIDCSTPCPLGT 447
Db 263 GYRGYCRPCRPAGFYGLCRRRCQKQOQCTVAEGRCILCEPGWNGTKCDPCATGF 322
Qy 448 YGINSRRC-GKNDVAVSPVDSCT--CKAGHGVDCSRGSGMVGSCNLTQCLNGS 505
Db 323 YEGGSHRCPPCRDHACHHVTGKTRCNAGNIGRCETKCSNGTYGEDCAFVCADCGSG 382
Qy 506 ACNTLDGTCTCAPGMRGEKCELPCQDGTGLNCAERCDSSHADGCHPTTGHC 557
Db 383 HDPSGRCICSPGVHGRHCNVTCTCPGLHAGDACACSC-HEDTDPTYGAC 433

```

Fri May 9 13:01:36 2003

us-10-092-390-2.rapb

Page 11

```

RESULT 15
US-09-900-449A-5
: Sequence 5, Application US/09900449A
: Publication No. US20030040616A1
: GENERAL INFORMATION:
: APPLICANT: ZHONG, Jenny et al.
: TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,
: TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND
: TITLE OF INVENTION: US$ THREEDE
: FILE REFERENCE: CL001271
: CURRENT APPLICATION NUMBER: US/09/900,449A
: CURRENT FILING DATE: 2001-07-09
: NUMBER OF SEQ ID NOS: 8
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 5
: LENGTH: 601
: TYPE: PRT
: ORGANISM: Strongylocentrotus purpuratus
: US-09-900-449A-5

```

Query Match	11.8%;	Score 795.5;	DB 9;	Length 601;
Best Local Similarity	30.5%;	Pred. NO. 6.6e-32;		
Matches 236;	Conservative 68;	Mismatches 211;	Indels 259;	Gaps 59;

QY	124	QCCEPFGMGITGSSACGDDHGFICTSRCCKNKALC-NTPTG-ACGCAAGPFGMRREDR	161
Db	16	CTCPFGFTGRNCLEIDID-----ECADP-FCQNGACVDDVNGVYCNVCYGFBDDECENN	68
QY	182	CEGTYGNDCHQRCOCONGATCDHVTG-----ECRCPRGYTGAFC--DLCPGKHGPOC	234
Db	69	I-----NEC-ASSPCLINGTIC--VDQVNMEECTCLAGFTGVRCENVIDEC-----	110
QY	235	EQRCPONGGVC-HHYVYG-EBCSPSGMGHTYVCQGPCEGFRGKNCSQEO-----CHNGTIC	289
Db	111	ASAPCQNGGCLIDINGITCSCPLFGSGDNC-----ENNDDECSITPLNGGTC	159
QY	290	-DAATG-QCHCSPGYTGERCO--DECPRVGTGYLCAETCCQVNGCKCYHVSAGCLCEAG	344
Db	160	VDLVNAYMYCAVPAMTGPITCADNIDEC-----ASAPCQNGG-----	195
QY	345	FAGRCEARLCPELTYGKCDKRCRPHLEMTNHSCHPMGSGBACKPGMSGLYC-----NETC	400
Db	196	-----VCIDVNGMYCD-----COPGTGTHCEIDTDECA	225
QY	401	SPFGYGEACQOICSCONGADC-DSVYTG-KCTCAPGFKGIDCSTPCPLGTGINKSSKGC	458
Db	226	RP-----PCQNGGPDVDVNGVYGCICAPGFGDNCENNID-----BDASR-PC	267
QY	459	KNDVACSPVDG-----SCTCAGMHGVDVCSIT--RCPSGTMGFGCNLTQCLNGACATLT-	510
Db	268	QNGVAVC-VBDVNVGFTVCTCSAGTYGVLCETDINECAS-----MPLNGGVCTDVLV	315
QY	511	DG-TCTCAPWRGSEKE-----LPCCDGTGLNLAERDCDSHADGCHPTTGH-CRCL	560
Db	316	NGYICTCAAGEBGINCELDIDECASFFCQNGA--TCTDQ-----VNGYCTCY	361
QY	561	PGMSGVHCD--SYCABGRMGPNCSLPCYCKKNASCSP--DDGICECABGFRGTTQORIC	615
Db	362	PGTYGVLCETDINECA-----SPCLCL-NGGTCDQVNGVYCAVCAODTSTVCTET--	409
QY	616	SPFGYGHRCQSQTCQCVHSSGPRCHITG--LCDDLPFTGALC--NEVPSRBRFGNCA	670
Db	410	-----DRDECASAPCL-NGGACADVVNGVYCTCLPMEGTNCEINTDECASS-----	455
QY	671	GICTNTNGTCNPID-----SCOCYPMWIGSDCSQPCRPAMWGNCLTHPCNCHNGAVC-	724
Db	456	---PCMNNGTLC--VDQVNSVYCFCLPFGTGHCTEIDECCASSP-----CLNGGGCID	503
QY	725	-SAYDDECKTTPMTGTYCQRRLPFGYKIDCALICQONGADC-DHISGO-CTCTGTG	780
Db	504	RVDST--ECVCAAGITVA--RCQIIN--DECA-SAPCQNGGCVDDVNGVYCNVCAPGY	554

```

Qy      781 MGRHCE---QKCPSGTYGYGCRQIDCLDNSTC-DHITG-TCYCSPEMKGARCD 829
       | :|| :| | || | : : | | | | : | -| :
Db      555 TGDNCETELDECAS-----MPLNGAGACIEWNGYTCCQCVAGTGVICE 598

```

Search completed: May 9, 2003, 12:02:17
Job time : 134.851 secs

```

OY 781 MGRHCE---OKCPSGTYGCGCRQICDCLNNSTC-DHITG-TCYCSPGMKGARCD 829
Db 555 TGDNCETFEIDECAS-----MPCLINGGACIEMWNGYTCQCVCAGITGVICE 598

```

Search completed: May 9, 2003, 12:02:17
Job time : 134.851 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 9, 2003, 11:53:35 ; Search time 22.4565 Seconds

(without alignments)
1493.646 Million cell updates/sec

Title: US-10-092-390-2

Perfect score: 6744
Sequence: 1 MVLINSLCFLICLLCHWI.....SSPKQEDSGSSSSSSSSSE 1140

Scoring table: BIOSUM62
Gapop 10.0, Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA:*

- 1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
- 2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
- 3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
- 4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
- 5: /cgn2_6/ptodata/1/1aa/PCFUS.COMB.pep:*
- 6: /cgn2_6/ptodata/1/1aa/backfill1a1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1037	15.4	2523	1 US-08-185-432-18	Sequence 18, Appl
2	1037	15.4	2523	4 US-08-899-232-3	Sequence 3, Appl
3	1035.5	15.4	2556	1 US-08-185-432-17	Sequence 17, Appl
4	1035.5	15.4	2556	4 US-08-899-232-2	Sequence 2, Appl
5	1034.5	15.3	2556	1 US-08-083-590A-20	Sequence 20, Appl
6	1034.5	15.3	2556	3 US-08-532-384-20	Sequence 20, Appl
7	1014.5	15.0	2471	1 US-08-185-432-16	Sequence 16, Appl
8	1014.5	15.0	2471	1 US-08-083-590A-19	Sequence 19, Appl
9	1014.5	15.0	2471	3 US-08-532-384-19	Sequence 19, Appl
10	1014.5	15.0	2471	4 US-08-899-232-1	Sequence 1, Appl
11	978.5	14.5	2703	1 US-08-185-432-19	Sequence 19, Appl
12	978.5	14.5	2703	4 US-08-899-232-4	Sequence 4, Appl
13	953.5	14.1	1964	4 US-09-467-997-1	Sequence 1, Appl
14	790	11.7	1055	4 US-09-214-278-2	Sequence 2, Appl
15	789	11.7	1212	4 US-09-214-278-3	Sequence 3, Appl
16	781	11.6	1257	3 US-08-611-729A-8	Sequence 8, Appl
17	779.5	11.6	1065	2 US-08-400-159-8	Sequence 8, Appl
18	779	11.6	1193	4 US-09-188-930-332	Sequence 332, App
19	779	11.6	1193	2 US-08-400-159-10	Sequence 10, Appl
20	779	11.6	1193	3 US-08-611-729A-10	Sequence 10, Appl
21	759	11.3	299	4 US-09-214-278-5	Sequence 5, Appl
22	759	11.3	299	4 US-09-188-930-192	Sequence 192, App
23	757	11.2	1036	4 US-09-068-740A-6	Sequence 6, Appl
24	757	11.2	1187	4 US-09-068-740A-7	Sequence 7, Appl
25	757	11.2	1208	4 US-09-199-865-1	Sequence 1, Appl
26	757	11.2	1218	2 US-08-400-159-6	Sequence 6, Appl
27	757	11.2	1218	3 US-08-611-729A-6	Sequence 6, Appl

28	757	11.2	1218	4 US-08-882-046-2	Sequence 2, Appl
29	757	11.2	1218	4 US-09-068-740A-11	Sequence 11, Appl
30	755.5	11.2	2199	5 PCT-US95-11684-2	Sequence 2, Appl
31	755	11.2	2199	4 US-08-882-046-7	Sequence 7, Appl
32	755	11.2	1148	4 US-08-882-046-4	Sequence 4, Appl
33	747	11.1	1218	4 US-09-214-278-7	Sequence 7, Appl
34	745	11.0	1248	4 US-08-882-046-6	Sequence 6, Appl
35	744	11.0	1219	4 US-08-882-046-5	Sequence 5, Appl
36	709.5	10.5	1810	5 PCT-US95-11684-4	Sequence 4, Appl
37	698	10.3	3075	2 US-08-460-309-5	Sequence 5, Appl
38	698	10.3	3075	2 US-08-125-077-5	Sequence 5, Appl
39	690	10.2	1404	2 US-08-400-159-2	Sequence 2, Appl
40	690	10.2	1404	3 US-08-611-729A-2	Sequence 2, Appl
41	683	10.1	3111	2 US-08-460-309-4	Sequence 4, Appl
42	683	10.1	3111	2 US-08-125-077-4	Sequence 4, Appl
43	623	9.2	1417	4 US-08-900-230-3	Sequence 3, Appl
44	595.5	8.8	2211	4 US-09-738-884-1	Sequence 1, Appl
45	585	8.7	1345	2 US-08-977-767-3	Sequence 3, Appl

ALIGNMENTS

RESULT 1
US-08-185-432-18
Sequence 18, Application US/08185432
Patent No. 5750652
GENERAL INFORMATION:
APPLICANT: Artavanis-Tsakonas, Spyridon
APPLICANT: Bussseau, Isabelle
APPLICANT: Diederich, Robert J.
APPLICANT: Xu, Tian
APPLICANT: Matsuno, Kenji
TITLE OF INVENTION: DETECT PROTEINS, NUCLEIC ACIDS, AND
TITLE OF INVENTION: ANTIBODIES, AND RELATED METHODS AND COMPOSITIONS
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/185,432
FILING DATE: 21-JAN-1994
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7326-006
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 2523 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-185-432-18
Query Match 15.4%; Score 1037; DB 1; Length 2523;
Best Local Similarity 25.4%; Pred. No. 1e-60;
Matches 326; Conservative 84; Mismatches 304; Indels 568; Gaps 78;
QY 83 GERTMYRR---KSCQ-----CP-GFYEGGECVPHCAKCVHGR 117

```

Db 53 GERCOPFNPCTIKNOCMNFGTCBPVLQNAIDFICHPVGF--TDKVCILPVDNACVNNP 110
  118 C-----IAPNTCOCEPBGWGTNCSSACDGDHMGPHCTSRQCKRGALCNP--IT 164
  111 CRNGTCELLNSTEXKCRCPGPMGTGDSQQA-----DPCASN--PCANGKCLPFEIO 162
  165 GACHCAFRGMRCE--DRCEQ-----GTGNDCHOR-----CQ 196
  163 YICKCPFRHGATCKODINCSNPKNGGQCINERFSGYCTQONFTGNCDPEYPCN 222
  197 ---CONGATC--DHVTGRCRPPGYTGAFCE-----LC 225
  223 PSPCLNGGTCRQTDTSYCTCLPFGSGQNCENIDCPSNNCRNGTGVGYNTYNCQ 282
  226 PRKHGPOCEQ---RC-----PCQNGVCHHTG--ECSOPSGMGTVCQ----- 266
  283 PDMTQOYCTEDVDECOLMPNACQNGGTCHNTYGYNCVNCWMTGEDSCSENIDCANNA 342
  267 -----PCEGRFRGKNC--SOEC---QCHNGTCD--ATGQ--CHCSPG 301
  343 CHSGATCHRVAIFYECCHGRGLLCHLDNACISNPNCSNCDINPNVNGKALCTCPG 402
  302 YIGERCQ---DECPVGTGYLCAETQCVNGKCYHVSQA--CLCEAGFAGERCEARLCP 356
  403 YTGPAKNNDVDECSLGN-----PCEHGRCTYTLGSPQNCPOGAGPRCEIDV-- 452
  357 EGLYIKCDKRC--PCHLEHTSCHPMSE--CAKPGMSGLYC----- 396
  453 -----NECLSNPC--QNDSTCLOIGEFQCIOMPGEGLYCEETINDECASNPCLN 501
  397 -----NE--TCSPGFYGEACQ---QICS---CONGADC-----DSYTK- 427
  502 GKCIDKINERCDCTPGFSGNLQHDPECTSTPCKNKAACLDGPNSTYTCQCTEGFTGRH 561
  428 -----CTCAPGFKGIDC-----STP----- 442
  562 CEODINECIIPDCHYGTCKDGLATFTYCLCPGTGRLCNDINECLSKPLNGGQCTDRE 621
  443 -----CPLTYGINSR--CG---CKNDVACSVDG--SCTKAGMHWVDCSIR----- 485
  622 NGYICCPGTTGVNCEKTRIDCASNLCDNGKIDKIDEGCECEGYGKLCININIEC 661
  486 -----CBSGTWFGFC----- 495
  682 DSNPCRNNGTCKDOINGFTCVCPDGYDHMCISEVNECSNPNICAHCHDGVNGYKDCDE 741
  496 -----NLTCQ---CLNGACNTLDGT--CTCAPGMREKEL-----PC-Q 530
  742 AGHSGSNCIDINNECESNPNOMNGTCKDMTGAYICTCKAGFSGPNCOJINIECSSNPLCN 801
  531 DGT-----YGLNC-----AERCD-----CSHADGCHPT--TGHCRLPGMS 564
  802 HGTGIDVAGYKNCMLPTGALCEAVLAPCAGSPCKNGRCKESDPEFSECECPGNG 861
  565 GVHCD-----SVCABGRMBPNSL-----PCYCKNGASC 593
  862 GGTCEJEDNECVNPRCNGATCQNTNGSYKCNCKPGYTGNCMDIDDCQPNCHNGGSG 921
  594 SPDDGI-----CECAPGFRGTGCR-----ICSPGFYGHRC 624
  922 S--DGINMFPCNPAFRPKCEDEDINECASNPCKNGANCTDCVNSTYTCQPGFSGIHC 979
  625 SOTCPQCVHSS--GPRCHITGL-----CDCLPGFTGALC--NE----- 658
  980 ESMTPTCTESSCENGTC--IDGINTFQCPGPGFTSGYSCQHDINCDKSPCLNGTQCD 1037
  659 -----VPSGRFRKNCAGI-----CTCTANNGT--NPIDRSQCPTPGATGSSCQD 702
  1038 SYGTGYKCTPOGTYGLNCLNLYRMCDSPCKNGKCMQNTNPFYR--CECKSGMTGYVCVY 1096
  703 ---CPPA--HMGPNCIHTONCHNGARC--SAVDGECCKTPGWTGLCTQRCPLGFGKNC 755

```

```

Db 1097 SVSCVAAKQGVDIYHL--CRNSGMCVDTGNTHPCRCQAGTYSYCEQV-----DEC 1148
  756 ALICQONGADC-DHISG-QCCTRTGFMGRHCBK----- 788
  1149 S-PNPOGATCTDYLGISCECVAGIHGVNCSSEETNECLSHPCONGGTCIDILINTYKS 1207
  789 CPSTGYGCRQICD-----CLNNTSC-DHITG--TCYCSPGMKARGCQDQAG 832
  1208 CPGRTGVHCEIIVDDCTPFYDSFTLEPKCFNNKGIKIDRAGVYNCICPFGVERGE----- 1264
  833 VIIVGNLS--LSRTSTALPADS 853
  1265 ---GDVNECLSN-----PCDS 1277

RESULT 2
US-08-899-232-3
; Sequence 3, Application US/08899232
; Patent No. 6436650
; GENERAL INFORMATION:
; APPLICANT: Qi, Huijin
; TITLE OF INVENTION: ACTIVATED FORMS OF NOTCH AND METHODS BASED THEREON
; FILE REFERENCE: 7326-046
; CURRENT APPLICATION NUMBER: US/08/899,232
; CURRENT FILING DATE: 1997-07-23
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 3
; LENGTH: 2523
; TYPE: PRT
; ORGANISM: Xenopus sp.
US-08-899-232-3

Query Match      15 4%; Score 1037; DB 4; Length 2523;
Best Local Similarity 25.4%; Pred. No. 1e-60; Indels 566; Gaps 78;
Matches 326; Conservative 84; Mismatches 304;

  83 GERTMYRR--KSQC-----CP-GFYESGMCVPHCADKCVHGR 117
  53 GERCOPFNPCTIKNOCMNFGTCBPVLQNAIDFICHPVGF--TDKVCILPVDNACVNNP 110
  118 C-----IAPNTCOCEPBGWGTNCSSACDGDHMGPHCTSRQCKRGALCNP--IT 164
  111 CRNGTCELLNSTEXKCRCPGPMGTGDSQQA-----DPCASN--PCANGKCLPFEIO 162
  165 GACHCAFRGMRCE--DRCEQ-----GTGNDCHOR-----CQ 196
  163 YICKCPFRHGATCKODINECSNPKNGGQCINERFSGYCTQONFTGNCDPEYPCN 222
  197 ---CONGATC--DHVTGRCRPPGYTGAFCE-----LC 225
  223 PSPCLNGGTCRQTDTSYCTCLPFGSGQNCENIDCPSNNCRNGTGVGYNTYNCQ 282
  226 PRKHGPOCEQ---RC-----PCQNGVCHHTG--ECSOPSGMGTVCQ----- 266
  283 PDMTQOYCTEDVDECOLMPNACQNGGTCHNTYGYNCVNCWMTGEDSCSENIDCANNA 342
  267 -----PCEGRFRGKNC--SOEC---QCHNGTCD--ATGQ--CHCSPG 301
  343 CHSGATCHRVAIFYECCHGRGLLCHLDNACISNPNCSNCDINPNVNGKALCTCPG 402
  302 YIGERCQ---DECPVGTGYLCAETQCVNGKCYHVSQA--CLCEAGFAGERCEARLCP 356
  403 YTGPAKNNDVDECSLGN-----PCEHGRCTYTLGSPQNCPOGAGPRCEIDV-- 452
  357 EGLYIKCDKRC--PCHLEHTSCHPMSE--CAKPGMSGLYC----- 396
  453 -----NECLSNPC--QNDSTCLOIGEFQCIOMPGEGLYCEETINDECASNPCLN 501
  397 -----NE--TCSPGFYGEACQ---QICS---CONGADC-----DSYTK- 427
  502 GKCIDKINERCDCTPGFSGNLQHDPECTSTPCKNKAACLDGPNSTYTCQCTEGFTGRH 561

```

```

QY 428 -----CTCAGPKGKIDC-----STP-----442
DB 562 CEGDINECIPDPCHYCTCKDGIATFTCLCRPGITGRLCNDINECSKFLNGOCTBRE 621
QY 443 -----CPLGTAVINCSSR---CG---CKNDVACSVPDC-SCCKAGMHGVDCSIR-----485
DB 622 NGYICCPGTTGVNCFETIIDDASMLCNGKCIDKIDIECECEGEGYTKLCNININC 681
QY 486 -----CPSGTWGFGC-----495
DB 682 DSNPCNNGSTCKDQINGFTCVCPDGYHDMCLSEVNECNSNPCIHGACHDGVNGYKDCDE 741
QY 496 -----NLTCQ---CLINGACNTLIDGT---CTCAGMWGKEKEL-----PC-Q 530
DB 742 AGHSGSNCDINNNECESNPMCMNGTCKDMTGATICTCKAGFSGPNCQTNINECSSNPCLN 801
QY 531 DGT-----YGLNC-----AERCD-----CSHADGCHPT---TGHCRCLPGMS 564
DB 802 HGTICIDVAGYKCNMLPYTGALCEAVLAPCAGSPCKNGRCKESEDFTFSECECPGMQ 861
QY 565 GHCD-----SYCABGRMCPNCSL-----PCYCKNGASC 593
DB 862 GGTCEIDMNECVNRCNATCQNTNGSYKCNCKPGYGRNCEMDIDCQPNPCNNGSGSC 921
QY 594 SPDDGI-----CECAPGFRGTTCOR-----ICSPGFYGHRC 624
DB 922 S--DGIMNFCNCPAGFRGPKCEEDINECASNPCNKGANCTDVSNTCTCOPFGSGLHC 979
QY 625 SOTCPGVHSS---GPRCHITGL---CDCLPFGTALC---NE-----658
DB 980 ESNTPCTESSCENGGTC---IDGIMFTCQCPGFGTGYCQHDINCDKSPCLNGTQCD 1037
QY 659 -----VPSRGEFKNCAGI---CTCTNNGTC---NPIDRSQCTPWTGLCTGRCLPGFSGSCQP 702
DB 1038 STGYCTCPGTYGTLNCLNLRWCDSSPCKNGKCMQTNNEYR-CECKSGMWGYCQDVP 1096
QY 703 ---CPRA---HMGPNCIHTNCNCHGAFRC---SAYDCECKCTPWTGLCTGRCLPGFSGSCQP 755
DB 1097 SVSCEFAAQQGVDIYHL---CRNSGMCVDTGNTHFRCAAGTGSICEQV-----DEC 1148
QY 756 ALICQCONAGDC-DHISG-QCTCRTGFMGRHCEQK-----788
DB 1149 S--PNPCONATCTDYLGSCECBVAGYHGVNCESEINECLSHPCQNGTCTCIDLINTYKCS 1207
QY 789 CSBGTGYGCRQICD-----CLNNTC-DHITG-TYCSGWMGARGARDQAG 832
DB 1208 CPBGTGVHCEINVDCTPFPYDSFTLEPRCFNNGKCIDRVGYNCTICPPGFVEGERCE--- 1264
QY 833 VIIVGNLNS-LSRTSTALPADS 853
DB 1265 ---GDVNECLSN-----PCDS 1277

```

US-08-185-432-17
Sequence 17, Application US/08185432

GENERAL INFORMATION:

APPLICANT: Artavanis-Tsakonas, Spyridon
APPLICANT: Bussieu, Isabelle
APPLICANT: Diederich, Robert J.
APPLICANT: Xu, Tian
APPLICANT: Matsuno, Kenji
TITLE OF INVENTION: DELTEX PROTEINS, NUCLEIC ACIDS, AND
TITLE OF INVENTION: ANTIBODIES, AND RELATED METHODS AND COMPOSITIONS
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.

```

ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/185,432
FILING DATE: 21-JAN-1994
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET INFORMATION:
TELEPHONE: (212) 869-9090
TELEFAX: (212) 869-8864/9741
TELECOMMUNICATION INFORMATION:
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 2556 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-185-432-17

```

```

Query Match 15.4%; Score 1035.5; DB 1; Length 2556;
Best Local Similarity 25.9%; Pred. No. 1.3e-60;
Matches 317; Conservative 84; Mismatches 304; Indels 519; Gaps 74;

QY 94 CCGPFESEGMCVPHCADRCVHGR-----IAPNTCCQEPGMSGTNGSSACDGDH 143
DB 89 CALGF--SGPLILTPLDNACLTPNCNNGSTCDLITLTKCKCPBMSKSOQA-----141
QY 144 WBPHTSRQCKNGALCNPTIGA--CHCAAGFRG--WRCEDRCGEG---TYGNDCHQ- 193
DB 142 ---DPCASN--PCANGGQCLPFEASYICHCPSEFGPTCMQDVNECGQKPRLCRHGTCHNE 198
QY 194 -----RC-----OCONATC---DHVGECCCPRGYTGARPE---222
DB 199 VGSYRCVCRATHTGPNCEMPYVPCSPSPCONGTCTPTGDTVTHECACLPFGTQNCENI 258
QY 223 DLCPRG--KHGPOC-----EORCP-----CONGVCHHTVS- 251
DB 259 DDCPGNCKNGACVADVNTYNCPCPEWYGTQCTEDVDEQCLMPACQNGGCTCHHTHG 318
QY 252 -ECSCPSGWMGYCQ-----PCPEGRFGKNC--SQEC-- 281
DB 319 YNCVGVNGWTGDECSENIDDCASACFHGATCHDRVASFYCBPCHGRTGLLCHLDACIS 378
QY 282 -QCHNGGTGDA--ATGQ--CHCSFGYTERCQ---DECFVGYGYVCAETCCQVNGKCY 333
DB 379 NPCNEGNSCDTPNPNVNGKAICTCPSGYTGACQDQDVDECSLGN-----PCEHAGKCI 430
QY 334 HTSQA--CLCEAGFAGERCEARLCPEGLYGKICDKRC--PCHLETHSCHPMSE--CA 386
DB 431 NTLGSFECCGLGYTPREIDV-----NECVSNPC--QNDATCILDQIGEFQCM 477
QY 387 CKPGWSGLYC-----NE--TCSPEFYGAQO---QICS--C 415
DB 478 CMPEGEVHCEVNTDECASSPCLHNGRCLDKINEFCCECPFTGHLQYDVEDCASRKY 537
QY 416 QNGADC-----DSV-TGKTCAPRFG 436
DB 538 KNGACLDLPNTYTCVTEGYGTGCEVDIDECDDPCHYSGCKGVATFTLCRPGYTG 597
QY 437 IDCST-----PCPL-----GYGINS-----SRCCKNDAYS 465
DB 598 HHCETININCSSQPCRLMWTGCDPDNAYLCFLKLTGTGNCENLDDDCASSPCDSGTCLD 657
QY 466 PYDG--SCCTCKAGMHGVDCSIR-----CPSGTWFGGCL-----TC--- 499
DB 658 KIDGYEACCEPGYTGSMCNSINDECAGNPNCHNGTCEIDGINFTCRCPGTYHDPICLSEV 717

```

```

OY 500 -----OCLNGACNTLDG-TCTCAPGMRGKCEL-----PCODGTGLNC 538
DB 718 NECSNPPCVHAGCWDLSNGTKCDDBGWSGTNCDINNNECESNPGVNGCTCKDMTSGIYC 777
OY 539 A-----ERC---DCSHADGC-HPTTGH-CRCLPGWSGVHDSV---CAEG- 575
DB 778 TCWEGFSGPNCOTNINECASNPCLNKGTCTIDVAGYKCNCLLPYTATCEVYLAPCAPSP 837
OY 576 -RMGPNC-----SLPVC---KNGASCSPDDGICECAPGFRGTTCORT-----CS 616
DB 838 CRNGGECROSEDESEFSVCVCPAGAKGTCEVDINECVLSPCMHSGASCONTHGXRYCHQ 897
OY 617 PGFYGHRSQTCPOC-----VHSSGPRCH--ITGLCDLPGFTGALCNE----- 658
DB 898 AGYSGRNCETDIDDCWPNRCHNGGSCDTGINTAFCDCLPGFTGCEEDINECASDPCRN 957
OY 659 -----VCPSGRFGKNCAG---ICT---CTNNGTCNPIDR---SCOCYPGWI 695
DB 958 GANCTDCVDSYTCCTPAGFSGICHENMTPDCTESSCFNGGTC--VDGINSFTCLCPGFT 1015
OY 696 GSDC-----SOP-----CPRAHGPNC---IHTCN---CHNGAFC 724
DB 1016 GSYCQHVNVECDSPRCLLGTCQDGRGLHRTCPQGYTGPNQNLVHWCDSPPCKNGKRC 1075
OY 725 ----SAYDGECKCTPGMTGLYCTQ-----R 745
DB 1076 WQTHQY--RCBPSGWTGLYCDVPSVCEVNAOROGVDVARLCOHGGLCVDAAGTTHCR 1133
OY 746 CPGLFYGKDCALI---CQ---CONGADC-DHISG-OCTCRFGMRHCEOK----- 788
DB 1134 COAGTGSYCEDLVDECSPPSPONGATCTDYLGYSCKVAGYHGVNCSSEIDECLSHPC 1193
OY 789 -----CPSGTYGCRQICD-----CLNNSGC-DHITG-T 816
DB 1194 ONGGTCLDLPMYTKSCSCMGVGVHCEINVDNCPVPDVPSNPKCFNNGTCTVDQVGGYS 1253
OY 817 CYCSPGKMGARCDQAGVIIVGNLN 840
DB 1254 CTCPPGFEVERCE-----GDVN 1270

```

```

RESULT 4
US-08-899-232-2
Sequence 2, Application US/08899232
Patent No. 6436650
GENERAL INFORMATION:
APPLICANT: Atavanis-Tsakonas, Spyridon
APPLICANT: Qi, Hulin
TITLE OF INVENTION: ACTIVATED FORMS OF NOTCH AND METHODS BASED THEREON
FILE REFERENCE: 7326-046
CURRENT APPLICATION NUMBER: US/08/899,232
CURRENT FILING DATE: 1997-07-23
NUMBER OF SEQ ID NOS: 4
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 2556
TYPE: PRT
ORGANISM: Homo sapiens
US-08-899-232-2

```

```

Query Match 15.44; Score 1035.5; DB 4; Length 2556;
Best Local Similarity 25.94; Pred. No. 1.3e-60;
Matches 317; Conservative 84; Mismatches 304; Indels 519; Gaps 74;

```

```

OY 94 CCGEYSEGEVCVPHCADKCVHGC-----IAPNTCCQEPGNGTNCSSACDGDH 143
DB 89 CALGEF--SGPLCTPLDNLACLTNCRNGTCDLTLTLEYRCRCPGWSGKSCQDA----- 141
OY 144 WGPHTSRCCCKNGALCNPTTGA--CHCAAGFNG--WRCEDEGEG--TYGNDCHQ- 193
DB 142 --DPCASN-PCANGGGLCPFEASTICHCPSPFHGPTCWDVNECGQKPRLCRHGGTCHNE 198

```

```

OY 194 -----RC-----OCONATC---DHVTEGCRCPGYTGAFCE--- 222
DB 199 VASRYCVRATHGTGPNCEMPYPSPSPCONGTCRPTGDTYHNECALPFGNGENBI 258
OY 223 DLCPRG--KHGPOC-----EQRCR-----CONGVCHHTVG- 251
DB 259 DDCPFGNCKNGACVDGVNFTYKPCRPBENTGQCTEDVDECOLMPNACONGTCHHTHG 318
OY 252 -ECSPSGMGTVCQ-----PCPEGRFGKNC--SOEC-- 281
DB 319 YNCVCVNGMTGDECDSEINIDCASAFHATCHDVAASYCCSPHGRIGLCHLDACIS 378
OY 282 -QCHNGGTDA--ANGQ--CHCSPGYGERCO--DECPVTTYGVLAETCCGVNGKCY 333
DB 379 NDCNNGSNDITPVPVNGKALCTCPSGYTGACSDVDECSIGAN-----PCEHAGKI 430
OY 334 HVSGA--CLCEAFGAGEREARLCEGLYIGICDRNC--PCHLENTSCHMSE--CA 386
DB 431 NTLGSFECQCLQGYTGPRCEIDY-----NECVSNPC--ONDAFTCLDQIGEFQM 477
OY 387 CKPWSGLYC-----NE---TCSPGFYGEACQ---QICS---C 415
DB 478 CMPGEGYHCEVNTDECASSPCLHNGRCLDKINEPQCCOPFTGHLQYDVBECASITPC 537
OY 416 ONGADC-----DSV-TGKCTCAPRFGK 436
DB 538 KNGAKCLDGPNTYTCVTEGYTGHCVEVDIDECDPDPCHYSCKKDGVATFTCLCPRYTG 597
OY 437 IDCST-----PCPL-----GTYGINS-----SRGCKNDVACS 465
DB 598 HHCEININECSSOPRCLMGTQODPNNATLCLCKTTPNCEINLDDCASSPDSGTCLD 657
OY 466 PVDG-SCITCKAGHGVDCSIR-----CPSGTWFGGNTL-----TC----- 499
DB 658 KIDGYECACEGYTSGMSNINIDECAGNPNCHNGGCEGINGFRCRCEGYHDPCLSEV 717
OY 500 -----OCLNGACNTLDG-TCTCAPGMRGKCEL-----PCODGTGLNC 538
DB 718 NECSNPPCVHAGCWDLSNGTKCDDBGWSGTNCDINNNECESNPGVNGCTCKDMTSGIYC 777
OY 539 A-----ERC---DCSHADGC-HPTTGH-CRCLPGWSGVHDSV---CAEG- 575
DB 778 TCWEGFSGPNCOTNINECASNPCLNKGTCTIDVAGYKCNCLLPYTATCEVYLAPCAPSP 837
OY 576 -RMGPNC-----SLPVC---KNGASCSPDDGICECAPGFRGTTCORT-----CS 616
DB 838 CRNGGECROSEDESEFSVCVCPAGAKGTCEVDINECVLSPCMHSGASCONTHGXRYCHQ 897
OY 617 PGFYGHRSQTCPOC-----VHSSGPRCH--ITGLCDLPGFTGALCNE----- 658
DB 898 AGYSGRNCETDIDDCWPNRCHNGGSCDTGINTAFCDCLPGFTGCEEDINECASDPCRN 957
OY 659 -----VCPSGRFGKNCAG---ICT---CTNNGTCNPIDR---SCOCYPGWI 695
DB 958 GANCTDCVDSYTCCTPAGFSGICHENMTPDCTESSCFNGGTC--VDGINSFTCLCPGFT 1015
OY 696 GSDC-----SOP-----CPRAHGPNC---IHTCN---CHNGAFC 724
DB 1016 GSYCQHVNVECDSPRCLLGTCQDGRGLHRTCPQGYTGPNQNLVHWCDSPPCKNGKRC 1075
OY 725 ----SAYDGECKCTPGMTGLYCTQ-----R 745
DB 1076 WQTHQY--RCBPSGWTGLYCDVPSVCEVNAOROGVDVARLCOHGGLCVDAAGTTHCR 1133
OY 746 CPGLFYGKDCALI---CQ---CONGADC-DHISG-OCTCRFGMRHCEOK----- 788
DB 1134 COAGTGSYCEDLVDECSPPSPONGATCTDYLGYSCKVAGYHGVNCSSEIDECLSHPC 1193
OY 789 -----CPSGTYGCRQICD-----CLNNSGC-DHITG-T 816
DB 1194 ONGGTCLDLPMYTKSCSCMGVGVHCEINVDNCPVPDVPSNPKCFNNGTCTVDQVGGYS 1253
OY 817 CYCSPGKMGARCDQAGVIIVGNLN 840

```



```

CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/532,384
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/083,590
FILING DATE: 25-JUN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Mastrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7326-015
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 790-9090
TELEFAX: 212 8698864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 2556 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-532-384-20

Query Match      15.3%      Score 1034.5: DB 3: Length 2556;
Best Local Similarity 25.8%      Pred. No.: 1.5e-60;
Matches 316: Conservative 83; Mismatches 304; Indels 523; Gaps 73;

94 CCGFYESEMCVPHCADKCVHGR-----IAPNTCCCEPGMGNGSSACDGH 143
89 CALGF--SGPLCLTPIDMNCITPCRNNGCTDLITLTKRCRPPMSKSCQDA----- 141
144 WCPHCTSRQCKNGALCNITGA--CHCAAGFERGRC--DRCEG--TYGNDCHO- 193
142 --DPCASN--PCANGGQCLPFEASVICHCPSEFGPTCRQDVNCGQKPRLCRHGTCHE 198
194 -----RC-----OCQNGATC--DHVTEGECRCPRGTYGAFCE-- 222
199 VGSYRCVCRATHTGNCERPRVYPCSPSPGONGTCRPTGVDVTHCACTLPFTGQNCENI 258
223 DLCPPG--KHGPC-----EQRCP-----CONGVCHHTG- 251
259 DDCPGNCKANGACVGVTYNCPPEMTGQCTEDVDECOQMPACQNGGCHHTHG 318
252 -ECSCPSGMGTVCQ-----PCPEGRFGKNC--SQEC-- 281
319 YNCVGVNGWTGECSEINIDCASAACFHGATCHDRVASFYCECPHRTGLCHLNDACIS 378
282 -QCHNGTDA--ATGQ--CHSPGTYGERQ--DECYVGYGVCAETGQCVNGSKCY 333
379 NPCNESNCDTNPVNCALCTCPSGTYGPACSDVDVDCSIGAN-----PCENHAKCI 430
334 HYSGA--CLCEAGFAGEBEREARLCPGLGIVGINKDKRC--PCHLENTSHCHPSGE--CA 386
431 NTLGSECCGLOGTYGPREIDV-----NECVSNPC--QNDATCLDIOIGFQCK 477
387 CRPGWSGLYC-----NE--TCSPGFYGRACQ--QICS--C 415
478 CMRGYGVHCEVNTDECASSPCLHNCRCLDKINEFCCECPGTGTHLQCYDVECASTPC 537
416 QNGADC-----DSV--TGKCTCAPGFKG 436

```

```

DB 538 KNGAKLDPEPNVTCVTEGTYGTHCEVDIDECDDPPCHYGSKDVAFTCLCRPGYTG 597
437 IDOST-----PCPL-----GYGINS-----SRCCKNAYVS 465
598 HHEETINECSSQPCRLRTGTCDDPDNAYLCFLCKGTGPNCEINLDDCASSPDSGTCLD 657
466 PYDG--SCTCKAGHWGVDCSIR-----CPSGTWGFGCNL-----TC----- 499
658 KIDGYCACPEPYTGSMKNSNIDECAGNPNCHNGTCEDDINFTCTCPRGYHPTCLSEV 717
500 -----QCLNGACNTLDG--TCTCAPGWRGKCEL----- 527
718 NECNSNPCHVAGACRDLNGLYKCDPCDPMGWTGNDINNNECESNPCVNGSTCKDMTSGIYC 777
528 PDDGTGYNLCAROD-----CSHADG--HPTTG--CHCLGWSGVHDSV--CAEG- 575
778 TCREGFSGRNCQTININECASNPCLNCTCIDVAGIKCKCLPRTGATCEVVLAPCAPSP 837
576 -RWGPN-----SLPCYC--KNGASCSPDDGICECAPGRGTTQRI-----CS 616
838 CRNGGECRGSDEYSEFSCVCPAGAKGTCCEVDINCEVLSPCRHGASQNTHGVRCHQ 897
617 PGFTGHRCSQTPQCVHSSGPRCH-----ITGLCDLPRFTALCNE----- 658
898 AGYSGRNCETDIDDC--RPNPCHNGSSCTDGINTAFCIDLPERGTFCEBINEDINECASDPC 955
659 -----VCPGSRFGKNCAG--ICT--CTNNGCNFIDR-----SCQCYPG 693
956 RNCANCTDCEVDYTCPCPAGFSIHCEENTPCTTESCNGNGTC--VDDINSFTCLCPRG 1013
694 WIGSDC-----SQP-----CPRAHWGPN--IHTCN--CHNCA 722
1014 FTGSGYOHVNECDSSRPLGLGTCQDGRGLHRCCTCPQGTGPNQNLVHMCSSPCKNGG 1073
723 FC-----SAYDGECKTRPMTGLXCTQ----- 744
1074 KCMQHTQY--RCECPSGWTGLXCDVPSVCEVAAROGVDVARILOHGCLVDAGNTHH 1131
745 -RCPLFGFKDCALI--CQ--CONGADC-DHISG-OCTCRGTFMGRICEOK----- 788
1132 CRQAGTYTSGYEDLVDEBSPSCQNGACTCTDYGISCKVAGYGVNCSBEIDCLSH 1191
789 -----CPSGTYGYGCRQICD-----CLNNTC-DHTTG 815
1192 PCQNGSTCLDLPTNYKSCSPRGTOGVHCEINVDNCPVDPVSRSEKCFNNGTCVDQVG 1251
816 -TCYCSPGWKGARCDQAGYIIVGNLN 840
1252 YSCTCPGFGVERCE-----GDVN 1270

RESULT 7
US-08-185-432-16
Sequence 16, Application US/08185432
Patent No. 5750652
GENERAL INFORMATION:
APPLICANT: Artavanis-Teakonas, Spyridon
APPLICANT: Bussseau, Isabelle
APPLICANT: Diedrich, Robert J.
APPLICANT: Xu, Tian
APPLICANT: Matsuno, Kenji
TITLE OF INVENTION: DETEEX PROTEINS, NUCLEIC ACIDS, AND
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:

```

```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/185,432
FILING DATE: 21-JAN-1994
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Mistock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7326-006
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 2471 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-185-432-16

```

```

Query Match          15.0%; Score 1014.5; DB 1; Length 2471;
Best Local Similarity 23.0%; Pred. No. 3,1e-59;
Matches 348; Conservative 106; Mismatches 354; Indels 707; Gaps 78;

```

```

93 QCCPGESEBMCVPHCADKCVHGR-CIAPNTCO-----CEFGWGTCSSACDG 141
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
91 RCASGF--TGEDCOYSTSHCFVSRPCLNGTCHMLSRDYECTCQYGFGRKEC----- 142
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
142 DHMGPHSTSCQCKNGALCPRTG--ACHCAGRGMRCDRCRQGYGGDCHQROCCN 199
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
143 -QMTDACLH-PCANGSTCTTANQFSCKLTGFTGKCE-----TDVNECDIPGHCH 194
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
200 GATCDHYTG--ECRCPRGYTGAFCEDL-----CPRGKHG 231
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
195 GGCCLNLPAGYGCQCGPQGFYGQYCDLSLYPCAPSPCVNGGTGQTGTFTECNLPRFEG 254
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
232 PCEBQ-----RCP-----CONGV 245
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
255 STGERNIDDCPNHRCONGVGVCDVGVNTYNCRCRPMWTGQFCTEDYDECLLPNACQNGGT 314
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
246 CHHYTG--ECSCPSGMGVTCGO-----PCEGGRFGKNC- 277
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
315 CARNRGYCYCVNMGSGDDCSENIDCAFASTPGSTCIDRYASFCMCPRGAGLLCH 374
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
278 -SQEC--QCHNGSTCDA--ATGO--CHCSPGYTGERCQ--DECPLYGYVLCAETCCQ 326
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
375 LDACISNPGCHKALCOTNPLNGYICTCPQGYKGAADCTEDYDECAM-----ANSNPC 427
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
327 VNGKCYHVGGA--CLCEAGFAGERCE-----ARLCPEGLY 360
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
428 EHAQKCVNTGAFHCECLKYGAFRCBMDINECHSDPCQDANCLDKIGFTCLCMGFK 487
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
361 GIKCDKR-----CP-----CHLE-----NTHSC 378
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
488 GVHCELEINECQSNPCYNNQCCYDKVNRFOCLCPRGFTGVCQIDIDDCSSTPLCLNACK 547
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
379 --HPMGECAKCPMGSLYCNET-----CSPGFGEAC-QQ 411
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
548 IDHNGYECOCATGFTVLCEBNIDNCDPDRCHHGQCCQDDISDYTCICNGYMGALISDQ 607
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
412 I-----CSCONGA-----DCDS--VTG-----K 427
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
608 IDECYSSPLCNDRCIDLVLNGYOCNCPGTSGVNCIEINFDCAENPCIHICMDGINRYS 667
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
428 CTGAPGKSIDC-----STPCPLGTGYNCS--RGCK-----NDAYCS 465
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
668 CVCSPGTGGRCNIDIDECASNPCKRKGATCINGVNGRCICLPBGPHPSCTSOYVNECLSN 727
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
466 P-VDGSCCT-----CKAGMHGVDCSI-----RCPSGT 490

```

```

DB 728 PCHHGNTGSLSGKCLCLDAGWGINCEVYKNECLSPCONGTGNTLVNGYRCTCKEP 787
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
491 WFGCNLTQC-----CLNGAC----- 507
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 788 KGYNCYINIDECASNPCNLNNGTGFDDISGYTCHVLPYTGKNCQTVLAPSPNCEAAV 847
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
491 WFGCNLTQC-----CLNGAC----- 507
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
508 -----NTLDTCTICAPGRMBREKCELPQODGTGLNCAERDCSHADCHFTTGH--CRCL 560
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 848 CKSPNESTYCLCAPGMQGRCTIDIDE-----CLSK-PCMHGKCHNTQGSYMECP 900
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
561 PGMSVHCDSDVCAEGRMGPMCSLPCYCKNGASCSPDDGI-----CECAPRGRTTCOR--- 613
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 901 PGSGMDCEIDDCLANP-----CONGSCM--DGVNTFSCCLIPGTGDKCQTDN 951
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
491 WFGCNLTQC-----CLNGAC----- 507
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
614 -----ICSPGYHRCSTQCPQCVHSS--GPCHHITGL--- 644
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 952 ECLSEPCKNGTGSDYVNSYTKCQAGFDGVHCENNINECTESSCFNGTGC--VDGINSF 1009
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
491 WFGCNLTQC-----CLNGAC----- 507
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
645 -CDCLPFTGALC-----NEV-----CPSGRGRKNC--AGICT- 674
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1010 SCLCPVGETGTFCLHETNECSSHPCLNETGTCVDGLTYRCSCLPGYGRKNCQTLVNLCSR 1069
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
491 WFGCNLTQC-----CLNGAC----- 507
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
675 --CTNNGTC--NPIDRSCQCYPMIGSDCSQP----- 702
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1070 SPCKNKGTCYQKKAESQCLCPSGMAGYCDVPNVSCDIASRRGVLEHLCOHSGVCINA 1129
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
491 WFGCNLTQC-----CLNGAC----- 507
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
703 -----CPRAHMGPNC--IHTC--NCHNGAFCSAYDG--ECKTPGTGLYCTQR-- 745
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1130 GNTHYCCPLGYTGSYCEBOLDPCASNPCOHGATCSDFIGYRCECPYGTGVNCEYEVD 1189
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
491 WFGCNLTQC-----CLNGAC----- 507
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
746 -----CPLEFGV-----KDCALICQCONGADG--DHISGO- 773
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1190 ECONQPCQNGTCLIDLVNHFKCSPRETGLLCEBNIDDCARGBHCLUNGQCHDRIGYS 1249
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
491 WFGCNLTQC-----CLNGAC----- 507
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
774 -----CTGRTGFMGRHCEQKPSGT 793
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1250 CRCLPFAGERCBGDIINECLSNPCSSBSGLDQIQLTNDYLCVORSAFTGRHCE-----T 1303
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
491 WFGCNLTQC-----CLNGAC----- 507
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
794 YGICRQICDCLNNSITCDHTTG-----TYCSPGMKARGCDQAGVILVNLNLSRSTRA 848
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1304 FVDVCPDM--PCLNGGTCVAVASNMPDGFICRCPGFSGFARQSS--CGGVKC----- 1351
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
491 WFGCNLTQC-----CLNGAC----- 507
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
849 LPADSVQIGAIAGIILVLFLLALFIYRHKQKKE-----SSMPAVYTTPAMRVVNA 904
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1352 -----RKGEQCVHTRASGR--CFCSPRDCES 1376
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
491 WFGCNLTQC-----CLNGAC----- 507
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
905 DYTISGTLPHSNGNANSHYFTNPSTHTLQCATSPHVNRRDRMTYKSKNNOLEFVNLKN 964
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1377 GCASS--PCQHGSGC--HPQRPPTYXS--CQA--PPFSGR-----CELYT 1415
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
491 WFGCNLTQC-----CLNGAC----- 507
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
965 VNGKRGVGDCTGTLPADMKHGGYLNELGAFGLDSYMKSLDKLAKNS--EYNSNC 1021
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1416 APPST--PPATYCLSYQCADARQVCDV-----ACNSHACQMDGDC 1455
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
491 WFGCNLTQC-----CLNGAC----- 507
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1022 SLSSSENPATIKDP 1036
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1456 SL-TMENPANCSSP 1469

```

```

RESULT 8
US-08-083-590A-19
; Sequence 19, Application US/08083590A
; Patent No. 5786158
; GENERAL INFORMATION:
; APPLICANT: Atavanis-Tsakonas, S. et al.
; TITLE OF INVENTION: Therapeutic And Diagnostic Methods
; TITLE OF INVENTION: And Compositions Based On No. 5786158ch Proteins And
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Penile & Edmonds
; STREET: 1155 Avenue of the Americas

```

```

CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/083.590A
FILING DATE: 25-JUN-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Mastro, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7326-015
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 790-9090
TELEFAX: 212 869864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 2471 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-083-590A-19

Query Match      15.0%; Score 1014.5; DB 1; Length 2471;
Best Local Similarity 23.0%; Pred. No. 3.1e-59;
Matches 348; Conservative 106; Mismatches 354; Indels 707; Gaps 78;

OY 93 OCCPFEYSGEMCVPHKADKCYHGR-CIAPNTCQ-----CEPGWGTCNGSACDG 141
DB 91 RASGF--TGECQYSTHPCFVSRLCLNGTCHMLSRFTYCTQVGTGKEC----- 142
OY 142 DHMGFHTSRCCCKNGALCNPTG--ACHCAGFRGWRCEDERCDEGTGNDCHOROCQN 199
DB 143 -QMTACLSH-PCANGSTCTVANQFSCKLTGFTGKCE-----TDVNECDIGHCQH 194
OY 200 GATCDHYTG--ECRCRPTGTGAFCDL-----CPRGKRG 231
DB 195 GGTCTANLPSTYCCCPQGCTGYCDLSLYPCAPSPCVNGTCRGTGDTFTFECNCLPFGEG 254
OY 233 PCOCE-----RCP-----CONGV 245
DB 255 SICEINIDDCRPHNRONGVYCDVNTYKRCRPPMTGQFCIEDYDECLDOPNACONGST 314
OY 246 CHHTVG--ECSCPQGMGTVCQ-----PCPGRFGKNC- 277
DB 315 CANRNGYGCVCVNGMSGDSCSENIDDCAFASCTPGSTCIDRVASFSCMPGKAGLLCH 374
OY 278 -SQEC--QCHNGGTCDA--ATGQ--CHCSPGYTGERQ--DECPYGYVLCATCQC 326
DB 375 LDDACISNPHCHGALCDTNPLNGQYITCTCPQGYKAGDCTEDVDECA-----ANSNC 427
OY 327 VNGKCYHVSQA--CLCEAGFAGERCE-----ARLPEGLY 360
DB 428 BHAGCYVMTDGAFHCECLKGYAGRPCEMDINECHSDPCONDATCLDKIGFTCLCMPRGK 487
OY 361 GIKCDKR-----CP-----CHLE-----NTHSC 378
DB 488 GVHCELEINECOSNPNVNGQCVDRVNRFOCLCPGFTGPVCQIDIDDCSSTPCINAGAC 547
OY 379 --HPMSGEACCKRGMGSLYCNCT-----CSPGYGKNC-QQ 411
DB 548 IDHPNGYECQATGFTGYLCEENINDCDBPCPHGQCQDCGIDSYCTICMPGYGALICSDQ 607
OY 412 I-----CSCONGA-----DCDS--VTG-----K 427
DB 608 IDECVSSPCLNDRGCLDLVNGYQCNQCPQTSVNCENINFDDCASNPCINHGICMDGINRS 667

```

```

OY 428 CTCABGFKGIDC-----STPCPLGTGYINCSS--RCGCK-----NDAYCS 465
DB 668 CVCSPGFTGQRNCINIDIDBCASNPCRKATCINGVNGFPCICEGPHHPSYQVNECLSN 727
OY 466 P-YDGSCT-----CKAGHGVDCST-----RCPST 490
DB 728 PCHNCNCTGSLGKYCLCDAGVNGVINCEDVKNELSNPCQNGTCDNLVNGYCTCKRGF 787
OY 491 WGFQCNLTQC-----CLNGGAC----- 507
DB 768 KGYNQVINDBCASNPCNLNGTCEFDISGYTCHCVLPTGKNCQVTLAPCSNPECNAAV 847
OY 508 -----NTLDGTCTCAPRNGERKCELPQDGTGYLNCARCDGSHADGCHPTTG--CRCL 560
DB 848 CKSPNFESYTLCAPMGOGKCTIDIDE-----CISK-PCMNHGLCHNTGGSYMCBP 900
OY 561 PGMVSYHCDYSVABEGRMPCNCLPCTCKNGASCSDDGI-----CCGATGFRITQOR--- 613
DB 901 PGFSGMDCEEDIDCLAND-----CONGSCM--DGVNPFSCICLPFGTGDCKQTDNM 951
OY 614 -----ICSPGYGHRCSQTCPCQVHSS-----GPCHHITGL--- 644
DB 952 ECLSEPCKNGTCSYVNSYTKCKQAGFDGVHCENNINECTPSSCFNGTGC--VGIINSF 1009
OY 645 -CDCLPFTGALC--NEV-----CPSGRFGKNC--AGICT- 674
DB 1010 SCLCPVGTGTSFCLHEINDCSSHPCLNBTGTVGLGTYRSCPLAGTKNCTQTLVNLCSR 1069
OY 675 --CTNNGTC--NPIDRSCQCYRGMVIGSDCSQ----- 702
DB 1070 SPCKNRKGTVCYOKKASQCLCPGSMAGAYCDVNVSCDIAASRGVLYEHLQHSQVCINA 1129
OY 703 -----CPRAHMGPNCC--IHTC--NCHNGAFCSAYDG--ECKTSPGMYLYTOR-- 745
DB 1130 GNTHYCQCPGLGTGSGYCEBQLEDCASNFCQHATGSDPTIGTRCVCYGYGVNCEYEVD 1189
OY 746 -----CPLGFTG-----KDALICQONGADC-DHISQ- 773
DB 1190 ECQNPQONGTCTIDLNVNFKCSPGTRGLLCEENIDDCARGPCLNGQCMDRIGYS 1249
OY 774 -----CTCRGTFGRHCEQKCPSTGT 793
DB 1230 CCLPGRFAGERCEGDIINECLSNPCSESSLDCTQLNDLYLCVCRAGFTGRICE-----T 1303
OY 794 YGYGCRQICDCLNNTCDHITG--TCYCSPGKRGARCDQAGVYIYGNLNSLSTSTA 848
DB 1304 FVDVCPQM--PCLNGGTCAVASNMPPGFTICRCPGRPSGARCOSS-----CGOYVC- 1351
OY 849 LPADSTQIGALAGIILVVLVFLALFLYRHKQKKE--SSMPAVTYYTPAMRVYNA 904
DB 1352 -----RKGEQCVHTASGPR-CFCSPRDES 1376
OY 905 DYTISGTLPHSMNGANSHYFTNPSTHTLLOCATSPHNVNNRBMVYTSKNNQJLVNLKN 964
DB 1377 GCASS--PCQGGSC--HPQRQPPYTS--CQCA-PPESGR-----CELYT 1415
OY 965 VNPGRGPFVGDCTGTLPAWDKHGGLYNEIGAFGLDRSYMGKSLKDLGRNS--EYNSNC 1021
DB 1416 APPST--PPATCLSQYCADKARDVGYDE-----ACNSHACQMDGDGQ 1455
OY 1022 SLSSENPYATITKDP 1036
DB 1456 SL-TWENPMWANCSSP 1469

```

```

RESULT 9
US-08-532-384-19
; Sequence 19, Application US/08532384
; Patent No. 6083904
; GENERAL INFORMATION:
; APPLICANT: Artavanis-Tsakonas, S. et al.
; TITLE OF INVENTION: Therapeutic And Diagnostic Methods

```


TITLE OF INVENTION: And Compositions Based On No. 6083904ch Proteins And
 TITLE OF INVENTION: Nucleic Acids
 NUMBER OF SEQUENCES: 21
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Pennele & Edmonds
 STREET: 1155 Avenue of the Americas
 CITY: New York
 STATE: New York
 COUNTRY: U.S.A.
 ZIP: 10036
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/532,384
 FILING DATE:
 CLASSIFICATION: 424
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/083,590
 FILING DATE: 25-JUN-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Mistrock, S. Leslie
 REGISTRATION NUMBER: 18,872
 REFERENCE/DOCKET NUMBER: 7326-015
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212 790-9090
 TELEFAX: 212 8698864/9741
 TELEX: 66141 PENNIE
 INFORMATION FOR SEQ ID NO: 19:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2471 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: unknown
 MOLECULE TYPE: peptide
 US-08-532-384-19

Query Match 15.08; Score 1014.5; DB 3; Length 2471;
 Best Local Similarity 23.08; Pred. No. 3.1e-59;
 Matches 348; Conservative 106; Mismatches 354; Indels 707; Gaps 78;

93 QCCGFESGEMCVPHCADKCVHGR-CIAPNTCO-----CERGMGTGNSACDG 141
 91 RCASGF--TGEDCOYSTSHPCFVSFRPLNGSTCHMLSRDYTECTQVGFTEKEC----- 142
 142 DHMGPHCTSRQCKNGALCNPTIG--ACHCAAGFRGWRCEDRBOGTYYGNDCHQRCOCN 199
 143 -QMTDACTLSH-PCANGSTCTTVANQFSCKCLTGFTGQKCE-----TDVNECDIPGHCH 194
 200 GATCDHWTG--ECRCPRGTGACFEDL-----CPPEKHG 231
 195 GGTCLNLPGSYQCCPGQFTGYCDLSLYPCAPSPCVNGGTGCTROTGTDETECNCLPFEBG 254
 232 PCEQ-----RCP-----CONGV 245
 255 STGRNIDDCPNHRCQNGVCYGVNTNCRCPQMTTGQCTEDVECLLQPNACQNGGT 314
 246 CHHTVG--ECCSPSGMGWTVCGO-----PCEGRFGKNC- 277
 315 CANNGYGVYCVNGMSGDCSENIIDCAFASCTPGSTCIDRYVASFSCMPEGKAGLLCH 374
 278 -SQCC--QCHNGTCA--ATGQ--CHGSPGTGERCO--DECPVGTGYVLCALFTCCQ 326
 375 LDDACISNPHKGLCTPLNPGYICTCPQGYKAGADCTEDVDECAM-----ANPNPC 427
 327 VNGKCYHVSQA--CLCEAGPAGERCE-----ARLCPGGLY 360
 428 EHAKCVNTDGAHFCECLKAGRCMDINECHSDPCQDATTCLKIGGFTCLCMDFK 487
 361 GIKCDKR-----CP-----CHLE-----NTHSC 378

488 GVHELEINEQSNPCVYNNQGVYKVRFCLOCPPTGTGVQCIDDDCSSTPOLNCAKC 547
 379 --HPMSECAKPRWGLYNET-----CSFGYGEAC-QQ 411
 548 IDHNGYECOCATGTGVLCEENIDNCDPDPCHHGQCCDIDSYTCICNPGYMALCISDQ 607
 412 I-----CSCONGA-----DCOS---VTG-----K 427
 608 IDECYSSPCLNDGRCLDLYNGVYQCNCPGTSGVNCCEINFDCCASNPCHIGLMDGILNRY 667
 428 CTCAPGFKGIDC-----STPCPLGTGYNCS--RCGCK-----NDAYCS 465
 668 CVSPFTGRCNIDIDECASNPCKRATLNGVNGRCLCPBPHHPCSYQVNECLSN 727
 466 P-VDGCT-----CRGWHGVDCST-----RCPST 490
 728 PCJHNGCTGGLSGYKCLDAGWVINCEDVKNNECLSNPCONGTCDNLVNGYRCTCKGF 787
 491 WFGCNLTQ-----CLNGGAC----- 507
 788 KGIYCOVINDECSNPCLNNGTCTPDDISGYTCHCVLPYTGKNCQTVLAPSPNCEMAV 847
 508 ---NTLDGTCYCAPGRKEKCELPQODGTGILNCAERDCSHADGCHPTTGH--CRCL 560
 848 CKESPNESTYCLCAPGMOGORCTIDIDE-----CISK-PCMNHGLCHNTQGSYMECP 900
 561 PGMSGVHCDVYCAGRGPRMCSLPCTYCKNGASCSPDGT-----CECAPRGRTTCOR-- 613
 901 PGFSGMDEBIDIDCLANP-----CONGSCS--DGVMTFSLCLPFTGTRKQCDMN 951
 614 -----ICSPGYVHRCSTQCPQCVHSS-----GPHCHITGL-- 644
 952 ECLSEPCKNNGTCSDIYVNSTYCKQAGFDVGHCENNINECTESSCFNGGTC--VDGINSF 1009
 645 -CDCLPFGTALC--NEW-----CPSGRPKNC--AGICT- 674
 1010 SCLCPVGTGTSFCLHEINECSHPCLENGTCVVDLGTGYRCSPLGYTGKNCQTLVNLCSR 1069
 675 --CTNNGT--NPIDRCOCYPRWIGSDCSOP----- 702
 1070 SPCKNKTQYQKAESECLCPSGWAGYCDVPNVSCDIASRRGVLEHLCHSHGYCINA 1129
 703 -----CPPAHMGPNC--IHTC--NCHNGARCSAYDG--ECKTPTGWTGLYCTOR-- 745
 1130 GNTHYCCPLAGTYSYCEBQDLBCASNPQHGATCSPTGIGRECYPRGYGVNCEYEVD 1189
 746 -----CPLEGY--KDCALICQCONGADC-DHISGO- 773
 1190 EQNQPCQNGTCLDLYNHFKSCSPPTRGILCEENIDCARGPHCLNGGQCMDRIGYS 1249
 774 -----CTGTGPMGRHCEQKCPSTG 793
 1250 CRCLPGRAGERCEGDINECLSNPCSSBSLDCIOLTDYDLYCVCSAFTGRCE-----T 1303
 794 YGGRQICDCLNSTDHITG-----TCYCSPPMKARCDQAGVILVGNLSLSTSTA 848
 1304 FVDYCPQM-PCNLNGITAVASNMPDGTICRCPGFSARQSS-----CGQYK- 1351
 849 LPADSYQIGALIGIILVVLLELLFTYRHKQKE-----SSMDAVTYTPAMRYVA 904
 1352 -----RKGEQCVHTASGR-CECPSPRCS 1376
 905 DYTISGLPHSNGNANSHYFTNPSTYTLQCATSPHYNNRDRMTYKSKNNQLFVULN 964
 1377 GCASS--PCOHGSSC--HROROPYYS-COCA-PPESGR-----CELYT 1415
 965 VNPGRGVPDCTGTLPADMKHGGVLELAFGLDRSYMKSJLDLGNK-----EYNSNC 1021
 1416 APST--PRATCLISQYADARQVDE-----ACNSHACQMDGGDC 1455
 1022 SLSSSENPTATIKP 1036
 1456 SL-TMENPMANCSP 1469


```

CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/185,432
FILING DATE: 21-JAN-1994
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Mistock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7326-006
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 2703 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-185-432-19

Query Match      14.5%; Score 978.5; DB 1; Length 2703;
Best Local Similarity 26.6%; Pred. No. 8,6e-57;
Matches 309; Conservative 105; Mismatches 331; Indels 417; Gaps 72;

QY 3 ISNSCLSTICLLCHWIG-----TASPLNLEDPNVC-SHWESYVTVQESYHPHD 53
DB 231 ICVNHGSIQCMCPGTGTRKDCDTKYKPCSPCQNAIGICRSNGLSYECKCPGEGKNC 290
QY 54 QIYITSCTDIL--NMFKCTRHRVSYRTAYRHGEKTYRRRSQCCPGFYESGEMC---VPH 108
DB 291 EONYDCCGLHLQNGSTCIDGISDY-----TCRCPPNF--TGRCCODDVE 334
QY 109 CADK---CVHGRCLAPNT-----CQCEPQNGVTC---GDHMGPHC-----T 149
DB 335 CAQRHPVQNG--ATCTNTHGSYSYICVNGMAGLDCSNNTDCKQNAACFYATCIDGVCS 393
QY 150 SPCOC---TNGALC-----NPTG--ACHCAGFFGMC---DRC 182
DB 394 FYCQCKGKGTGLCHLDACISNPHADALCDTSPINGSYACSCATGYRGVDCSEDIDEC 453
QY 183 EGGTGNDCRQCCONGATCDHVTG--ECRCPPGYTGAFCED----- 223
DB 454 DQGS-----PCEHNICVNTPSYRNCNSQSGFTGRCCEININECSHPCQNGSCL 504
QY 224 -----LCPPKHPGCE-----QRCPCQNGVCH--HYTG--ECSCEPSGMGTVC--- 264
DB 505 DDPGRFCVCMGFTGTQCEIDIDECQSNPCLNDGTCHDKINGKSCALGFTGARCQIN 564
QY 265 -----GCP-----CPGRFGKNC---QECQ---CHNGTCDATG-Q 295
DB 565 IDDCOSQPCRNNGICHDSIAGYSCECPRGYTGTSCEININDSDNCHGKCIDVNSFK 624
QY 296 CHCSPGTGERCO--DECP-----VGTGYVIC-----AETCO----- 325
DB 625 CLCDPRYTYICOKOINBESNPCOFDGHQDRVGSYTCQCAQAGTSGKNCVAVNECHSN 684
QY 326 -CVNGSKCYHVSQ---ACLCEGAFGERCEAR-----LCPEGLYGIKCDKR 367
DB 685 PNNNGATC--IDGINSYKQCVGFTGCKEKVDECISSPCANNVCIDQVNGYKCE-- 740
QY 368 CQCHLENTH-----SCHPMSEGCACKRPMGSLCNETCSRGFTGEACQ---QICS-- 414
DB 741 CPFGYDAHCLSDVDECAASNPCVNEGRCEGIDNEFTCH--CPGTYGKRCCELDIDECSSN 798

```

```

QY 415 -CONGADC-DSVYG--KCTCAPGFKGIDCST-----PCPIG----- 446
DB 799 PCOHGTCYDKLNAAFSCQCMPTGTGOKCETINDDCTVINCNGNGTIDKYNKCYKVP 858
QY 447 TYGNCSSR---CG---CKNDVCSF-----VDGCTCKAGMHGVDCSIRCPSGTWGFCN 496
DB 859 FGRDCEKMDPCARRNCKNEAKCPSPSNFLDFSCYCKLGYGRYDEDE-----CS 912
QY 497 LTRCCLNGACQNTLDST--CTCAPGRGKCEL-----POQDFTYGLNACRDCSH 546
DB 913 LSSPCRNASCLNVPSTYGLCTKYGKRDCAINTDDCASFPQONR-----TCLD 963
QY 547 ADGCHPTTGHCRCLPMSGVHDSVCAEGRMGPN--SLPCYCKNAGASCSP--DDGICECA 603
DB 964 GIGDY-----SCLCYGFDCKHCETDNE-----CLSQP--CONGATCQYVNSTYCTCP 1011
QY 604 PGFRGTTQR-----ICSPGFYGRCSOTPCQV--HSSGPCHHITGLDC 647
DB 1012 LQFSGINCQTNDEDTRESSCLNGGSCIDGINGNCS-----CLAGSGANCOYKLNKCD 1066
QY 648 LQFGLALCNEY-----CPSGRFGKNC-----GICTCNNGTCNPIDR--SCQCP 692
DB 1067 NPCLNATCHEQNNNEYTCHCPSPGFTGKCSEYVDMCGOSPCENGATCSQMKHOFSCGSA 1126
QY 693 GWIGSDCSQPCPAHMGPNCHITCNCH-----NAGFCSAYDGE--CKCTPG 736
DB 1127 GWTKRCLD-----VOTISQDADARKGLSLRQLCNNGTCKDKDGNSHVCCSGG 1174
QY 737 WTGLICTORCPLEFYKDCALICQ---CONGADCHISG--OCTCRFGMRHCEQ----- 787
DB 1175 YAGSYCKEID-----ECQSOPCONGCTCRDLIGAYECQCRQGOGNCLEINID 1224
QY 788 -----KPSGTGYGC---RQID--CLNNSC-DHITG--TCY 818
DB 1225 CAPNCPQNGCTCHDRYAMFSCSPETMGTICEINIKDKCPGACHNNGSCIDRVGSECV 1284
QY 819 CSPGMKARCDQAGVLIIGNLN 840
DB 1285 CQPGFVARGC-----GDIN 1299

RESULT 12
US-08-899-232-4
; Sequence 4, Application US/08899232
; Patent No. 6436650
; GENERAL INFORMATION:
; APPLICANT: Artavanis-Tsakonas, Spyridon
; TITLE OF INVENTION: ACTIVATED FORMS OF NOTCH AND METHODS BASED THEREON
; FILE REFERENCE: 7326-046
; CURRENT APPLICATION NUMBER: US/08/899,232
; CURRENT FILING DATE: 1997-07-23
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 4
; LENGTH: 2703
; TYPE: prt
; ORGANISM: Drosophila sp.
US-08-899-232-4

Query Match      14.5%; Score 978.5; DB 4; Length 2703;
Best Local Similarity 26.6%; Pred. No. 8,6e-57;
Matches 309; Conservative 105; Mismatches 331; Indels 417; Gaps 72;

QY 3 ISNSCLSTICLLCHWIG-----TASPLNLEDPNVC-SHWESYVTVQESYHPHD 53
DB 231 ICVNHGSIQCMCPGTGTRKDCDTKYKPCSPCQNAIGICRSNGLSYECKCPGEGKNC 290
QY 54 QIYITSCTDIL--NMFKCTRHRVSYRTAYRHGEKTYRRRSQCCPGFYESGEMC---VPH 108
DB 291 EONYDCCGLHLQNGSTCIDGISDY-----TCRCPPNF--TGRCCODDVE 334
QY 109 CADK---CVHGRCLAPNT-----CQCEPQNGVTC---GDHMGPHC-----T 149

```

```

Db 335 CAORDHPVQNG-ATCTNTHGSYSYICVGNMAGLDCSNNTDCKOACACATGATCIGVGS 393
Oy 150 SHQOC---KNGALC-----NPTTG---ACHCAAGFGWRC---DRC 182
Db 394 FTQCCKTKGTGLCHLDACSTNPFCHADAICDTPSINGSYACSCAAGYGVCSSEIDEC 453
Oy 183 EOGTYGNDCHORCOCCNGATCDHVTG--ECRCPPGYTGAFCED----- 223
Db 454 DQGS-----PCENHCICVNTPSTSYKNCNSQGTGTRCETNINECSHPCQNEGSL 504
Oy 224 -----LCPPKKGPOCE-----QRCPCQNGVCH-HVTG-ECSCPSGMWGTVC--- 264
Db 505 DDPTGFRVCMPGFTGTCEIDIDEQCSNPLCNDGCHDKINGFKSCALGFTGANCOQLN 564
Oy 265 -----GDP-----CPBGRFKNC--QEQO---CHNGTCDAAATG-Q 295
Db 565 IDDCSQPCRNAGICHDSIAGYSCBPPGYTGTSCENININDCSNPHKNGKICIDVNSPK 624
Oy 236 CHCSFGYTGRCQ---DECP-----VGTGVLC-----AETCQ----- 325
Db 625 CLCDDGYTGXICQKQINECSNPFQFDHCQDRVGSYCCQCAAGTSGKNCCEVNVNDECHSN 684
Oy 336 -CVNGKCYHVSQ---ACLEAGFAGERCAR-----LCPEGLYGIKCDKR 367
Db 685 PCNNATC--IDGINSYKCOCPVFTGQHCEKNVDECISPPCANNGVCIDQVNGYKCE-- 740
Oy 368 CPCHLENTH-----SCHPMSECBACKPGMSGLXCMETCSRGTYGACQ---QICS-- 414
Db 741 CPRGTYDAHCLSDVDECSNPFVNEGRCEDEGINETICH--CPPGYTGKRCLELDIDECSSN 798
Oy 415 -CQNGADC-DSVTG-KCTCARGFKGIDCST-----PCPLG----- 446
Db 799 PCOHGCTCYDKLNAFSCQCMPTGYTQKCEITNIDCVTPNPGNGGTCIDKVNKYKCYKVP 858
Oy 447 TYGINCSR---CG---CKNAVCSF---VDGCTCAGMHGVDCTSRCSGTWGFQCN 496
Db 859 FTGRCCSKMPCARNRKNKAKCTPSSNFTDFSCCTKLGTYGRYCDIDIDE-----CS 912
Oy 497 LTCOCLNGACNTLDGT--CTCARGMREKCEL-----PCOOGTGLMCAERCDSDH 546
Db 913 LSSPCRNAGASCLNPGSVRCICTKGYBEGRDAINTDDCASTPCQNGR-----TCLD 963
Oy 547 ADGCHPTTGHRCLEPGMSGVHCDVCAEGRMPCN-SLPCYCKNGASCSP--DDGICECA 603
Db 964 GIGDY---SCLCYDGFDEKHCETDINE-----CLSQP--CQNGATCSQYVNSTCTCP 1011
Oy 604 PGFRTTCOR-----ICSPGFYGHRCSTQCPQCV--HSSGCPHHITGLCDC 647
Db 1012 LGFSGINCQTNDECTESSCLNGSSCIDGTINGNCS-----CLAGYSAGANCOYKLNKDS 1066
Oy 648 LPGFTGALCNEV-----CPSGRFRKNC---GICTCTNNGTCNPIDR--SCOCYP 692
Db 1067 NPLCANATGCHBQNNNEYTCHGSSFTGKOCSEYVWMCQSPENAKTSQMKHOFSCKSA 1126
Oy 693 GWIGSDCSQCPRAHMGPNCTHTCNCH-----NGAFCSAYDGE--CKCTPG 736
Db 1127 GWTGKCLD-----VQITISCODAADRKGLSLRLCLNNGTCKDYGSHVYCYSQG 1174
Oy 737 WTLGYTORCPLGFGYKDCALICQ---CONGADCDHISG--QCTCRFGPMRHEQ--- 787
Db 1175 YASSTCKEID-----ECOSQPCQNGGTCRDLIGAYECOCRGFGQONELIID 1224
Oy 788 -----KCPSTGYGVC---ROICD---CLNNSTG-DHTTG--TCY 818
Db 1225 CARNPCQNGTCHDRVNVNPFSCSCPRTGWTGILCEINKDDCAFGACHNNGSIDRVGGECEV 1284
Oy 819 CSFGMKGARCDAQVITVGNLN 840
Db 1285 CQPFVAGARCE-----GDIN 1299

```

RESULT 13

```

US-09-467-997-1
; Sequence 1, Application US/09467997
; Patent No. 6379925
; GENERAL INFORMATION:
; APPLICANT: Kitajewski, Jan
; TITLE OF INVENTION: ANGIOGENIC MODULATION BY NOTCH SIGNAL TRANSDUCTION
; FILE REFERENCE: 53863-A-PCT-US
; CURRENT APPLICATION NUMBER: US/09/467,997
; CURRENT FILING DATE: 1999-12-20
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1964
; TYPE: PRT
; ORGANISM: mouse
US-09-467-997-1

Query Match      14.1%; Score 953.5; DB 4; Length 1964;
Best Local Similarity 25.0%; Pred. No. 2.8e-55;
Matches 297; Conservative 75; Mismatches 303; Indels 513; Gaps 61.

Oy 109 CADKCVHGRCI-----APNTCOCEPGMGNGCSACDGDHMGHCHTSROCKNGALC----- 160
Db 32 CAN-----GGTCRLRSNGGIGICQCAPFLGETC-----QFPDPCRDOLCKNGSGCAL 81
Oy 161 -----NPTTG--ACHCAAGFRGNRCEDR----- 181
Db 82 PTPPSSRSPTSLPTPHFSCOTCPSTGTRDRCOTHEBELCPSTSCNGHCYVAGSRPQCS 141
Oy 182 CEQGYTGNDCHOR--CQ---CONGATC--DHVTGCRCPPTGYTGAFCED----- 223
Db 142 CEPGWTGQCOLRDFCSANPANGVCLATYPOICRCRPPGFEGTTCERDINECFLEBP 201
Oy 224 -----LCPPKKGPOCEQR---CP---CONGVY-----HVTGGEQSC 255
Db 202 CPGGTSCHNTLGSVOCLCPVQOEBQCKLRGACPPSGCLNGGTQLYPEGHSTFHLCLC 261
Oy 256 PSGMNGTYCGO-----PCPEGRFRKNCQO---ECO----- 282
Db 262 PPGFTGLDCEMNPDCVHRHOCNGATCLDGLTYTCCPCKTWKMGDCSEDDIDECEARGP 321
Oy 283 -CHNGGTCDAAATG--OCHCSPGYTGRCODE-----CPVG 314
Db 322 RCRRNGTQNTAGSFHCYCVSGMGAGCEENLDDCAATCAPGTCIDRVGSFSCCLCPG 381
Oy 315 TYGVLC-----AETCQ----- 325
Db 382 RTGLCHLEDBCLSPCHVNAQSTNPLTGSTLICQPGYSGSYCHODLDCQMAOQPS 441
Oy 326 -CVNGKCYHVSQA--CLCEAGFAGERCAR-----LCPEG 358
Db 442 PCEHGSSCINTPPGSFNCLCIPGYTGSRCEDAHNECLSQPHGPGSTCLDLATFHLCLPPG 501
Oy 359 LYGIKCD---KRC---PCHELENTHSCHPMGS--EACKPGMSGLXCM----- 398
Db 502 LBEGLCEVNEECTSNPC--LNOACHDLNLNGFOCLDLPGETGARGCKRDECSSTPCAN 559
Oy 399 -----TCSBGFYGEACQOICS-----CONGADCSVTGK--CTCARGFGKI 437
Db 560 GGRCRDQPAFYCEBLPGFSPHCEKEVDBELSDPCGVASCLDLPALFAFLCLCRPFTGQ 619
Oy 438 DCSFP-----CPLYGTI-----NCSRSG----- 457
Db 620 LCEVPLCTPNMCOPOGQOCQOEHAPCLCPDGSPGCVAPAEKDPCHHGHORSLCYCDEG 679
Oy 458 -----CKNDVCSPYDG--SCTCKRGMHGVDCS---TICPSGTW--- 491
Db 680 WTPBECTELGCGISTPCAHGATGCHPOPSGYNCTCPAGYKGLTCSSEVNTACHSGPCLNG 739
Oy 492 -----GFGCN-----LTCOCLNGACNTLDGT--CTCARGMREK 525
Db 740 SCSTIREGYCTCLPSTHGHCTAVDHCYASASCLNGGTGVNRPRTFFCLCATGFGGLHC 799

```

```

Oy 526 E-----LPQDGTGLNC-----AERDQSHADGCHPT 553
Db 800 EKPTRSCADSPCKRKAATCQDTPRGARCLCSPGYTGSSCQTLIDLCARPCPHTRACLOS 859
Oy 554 --TGHRCCLPGMSVHCD--SYCABGRWGPNCSLPCYCKNGASCSPDDG--ICECARGF 606
Db 860 GPSFQCLDQWGTALCDPFLSCQKAAASQGIIEISGLCQNGSLCI-DWSSYFCRCPPGF 918
Oy 607 RGTTCORICSPGFYGHRCRSQTCPOCVHSSGCHHITGLCDLPGTGTALCNEVCPSGRRG 666
Db 919 QGKLCQDWNVP-----C--EPNPGHGS--TCVPOPSGYVCQ--CAPGYEG 958
Oy 667 KNCAGI---C---TCTNNGTC--NPIDRSOCQYPMIGSDC-----SQCPRAHNGP 710
Db 959 QNCSVYLADACOSQPCNHGHTCTSRPGFHCACPGFVGLRGGDVDECLDRCPHPS---- 1014
Oy 711 NCIHTCNGHAFCSAYDGECKTPGWTGLYC----- 742
Db 1015 ---GTAACH--SLANAF--YQOCLRGHTGQRCVEYEMDLCQSQPCSNSSCETITGPPGF 1067
Oy 743 TORCPLGYGKDC---ALIC---OCQNGADC-----DHSQCTCRGPMGRHC-EQKCP 790
Db 1068 TCHCKGEGEPTCSHKALSCGIIHCHNGGLCLPSRKPSPPLACLSGFGGPDCLTPPAP 1127
Oy 791 SGTVEYGGRCQICDCLNNTCDHITG-----TCVSPGMKARGCDQAG 832
Db 1128 P-----GGPPSPCLHNGTCTETPELGNPFGQCTCPPSPGRCRCPG 1170

```

```

RESULT 14
US-09-214-278-2
; Sequence 2, Application US/09214278
; Patent No. 6291210
; GENERAL INFORMATION:
; APPLICANT: Sakano, Sei11
; TITLE OF INVENTION: DIFFERENTIATION-SUPPRESSIVE POLYPEPTIDE
; FILE REFERENCE: KP-8576
; CURRENT APPLICATION NUMBER: US/09/214,278
; CURRENT FILING DATE: 1999-01-26
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Patentia Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1055
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-214-278-2

```

Query Match 11.7%; Score 790; DB 4; Length 1055;
 Best Local Similarity 27.7%; Pred. No. 1e-44;
 Matches 291; Conservative 85; Mismatches 377; Indels 296; Gaps 69;

```

Oy 15 LICHNIGTASPLNLEDPNVCSHMESYSTVVOESPYPHDPQIYYSCTFLMFKCTRHRY 74
Db 135 LIIERVSHNAGMINPEDRKMSLHFSGHVAHLEQIVRCDENYSTATCN-----KCRPRN 189
Oy 75 SYRTAYRIGEKTMRRKSOCCGFESEGBMCPH-CADKC--VHGRCTAPNTQCEPFG 131
Db 190 DF--FGHYTCDOYGNK-CMDGM--MGECKEAVCKOGCNLHGGCTVPGRCRSYQW 243
Oy 132 GTCSSACDGDHMGPHCTSRCKNGALCNPTGACHCAAGRGMRCEDRGQIYNDG 191
Db 244 GRFCD-----ECYPYRGCYVSGVEP--WQCNCEITMNGLLCDL-----NYC 285
Oy 192 HORCOQNGATCDHYTGE--CRCPGYTGAFCEDLCPGKHGPOCEORC--PCQNGV 245
Db 286 ESHNCTNGCTCINAEPPDQYRCTCPDGYSGRNC-----KAENHCTSNPCANGS 335
Oy 246 CHHYTG--ECSPSGMNGVC-----GQP-----CEPGRGKNC- 277
Db 336 CHEVPSGFECHCPSGMSGPTCALDIDECASNPCAAGTICVDYDGFECTICPQWNGATCQ 395

```

```

Oy 278 --SOECQ---CHNGTCDAAATG--OCHSPGYTGERCO---DECPVGTGYVLAEETCCQV 327
Db 396 LDANCBEGKPCPLNARFSCKNLIGYYCCDCLPGKKGINCHINYNVC-----RGCCQ 444
Oy 328 NGGKCYH-VSG-ACICEAGFAGECEARLCPEGLVIGIKDRKPCPLENTHSCHPMSEGC 385
Db 445 HGGCTKLDVNGYQYCCPGRGGRHCE-----LERDKC-ASSPCH-----SGG- 485
Oy 366 ACKPMSGLYCKNETSPGFYGEACQ--QICS---CQNGADCSTYTK--CTCAGFGGI 437
Db 486 LCEDLADGFHCH--CPOGFSGPLCEVDVLDLCPSPCRNGARCYNIEGDYCCAPDDFGK 543
Oy 438 DCST--PCPLGTGYINCSSRCCKNDA-----VCSPYVSGCTCKAGMIGVDCS 483
Db 544 NCSYREPCPGGA-----CRVIDGCSDDAGPGRHGTAAAGVCGP-HGRVYSQPG--NFS 595
Oy 484 IRCPSGTWGFQCN-----LTCQCLNGAGC-NTLDG-TCTCAPGRGECCE-----LP- 528
Db 596 CIGDSGFTGYCHENIDOLGQPCRNNGGTCTIDEVDAFRCFCPSGMEGELCDTNPDCLPD 655
Oy 529 -----CQDGTGLNCAER--CD---CSHADGCHPT--TGHRCCLPG 562
Db 656 PCHSRGRCYDLYNDFYCACDDGMKRTCHSRFQCDAYTCSNGTGYDSGDTFRACAPPG 715
Oy 563 WSGVHCDSYCAABGRWGPNCSLPCYCKNGASC--SPDDGICECAPGFRGTQORICSPGY 620
Db 716 WKG---STCAVAK--NSSC-LPNPCVNGGTGYGSAASFTICRDMEGRTCT----- 761
Oy 621 GHRCSQTCPOCVHSSGPHHITGL-----CDCLPFTGALCN--EVCPSGRGKNAGIC 673
Db 762 -HNTDNCNPLPCYNIGIC--VDGVNMFRECAPAGFPCCRINIDECOS---SPCAYGA 814
Oy 674 TCTN--NGTCNPIDRSOCQYPMIGSDC-----SQCPRAHMGKPCNIHCNCN 720
Db 815 TCVDEING-----YRCSPPEGRAPRCQEVYIFGRSCWSRTPPPH-GSSVVECCNS-- 865
Oy 721 GAFCSAYDGECKTPGWTGLYCTORCPLGFYKDCALICQCONGADC-DHISGOC---TC 776
Db 866 ---CCLDGRRCRCSKYWG--WKPCLLA--QOPALSAQCLGRCLEKAPGQCLRPCC 917
Oy 777 RT-GFMGRHCEQKCP---SGTYGRCQICDCLNNTCDHI---TGTCCYS-----PG 822
Db 918 EAMGECGAEEPTSTCLPRSHLDNNCARLTJLHFN---DHVPGQTYGALCSGIRSLPA 974
Oy 823 WKGARCDQAGVITVGNLNS-LSRTSTA-----LPADSYIGALAGIILVLYVPL 873
Db 975 TRAVARDRLVLLCDRASASAVEAVSFPARDLPSSLIQGAHALVAI----- 1027
Oy 874 ALFIYRHOKGKESMPAVTYTPAMRYV 902
Db 1028 -----TQKGNSSLLAIVTEVKYETV 1048

```

```

RESULT 15
US-09-214-278-3
; Sequence 3, Application US/09214278
; Patent No. 6291210
; GENERAL INFORMATION:
; APPLICANT: Sakano, Sei11
; TITLE OF INVENTION: DIFFERENTIATION-SUPPRESSIVE POLYPEPTIDE
; FILE REFERENCE: KP-8576
; CURRENT APPLICATION NUMBER: US/09/214,278
; CURRENT FILING DATE: 1999-01-26
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Patentia Ver. 2.1
; SEQ ID NO 3
; LENGTH: 1212
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-214-278-3
Query Match 11.7%; Score 789; DB 4; Length 1212;

```

Best Local Similarity: 26.8%; Pred. No. 1,4e-44;
Matches 309; Conservative 93; Mismatches 417; Indels 332; Gaps 74;

```

QY 15 LCHWMTGAPLNEEDPNVSHWESVTVQESVPHFDIYITSCDILIMFECTHRY 74
DB 135 LLERYSHAMINPEDMKLHFSGVHAIHELQIRVNCDEYISATCN-----KECPRN 189
QY 75 SYRTAVRHGEKTYWRKRSQCCPGFESEGCVPB-CADKC--VHGRCIAPNTCCCEBGMG 131
DB 190 DF---FGHYTCDDYGNKA-CMDGW--WGRCKEAVCKQGNLHGGCTVGECCSYGMQ 243
QY 132 GTNCSSACDDDHGPHCTSCCKNGALCNPTIGACCAAGFKGRKEDRCDEGTGNDG 191
DB 244 GRECD-----ECVPRGCVHGSCEVP--WQCNCEITNMGSLCDKDL-----NYC 285
QY 192 HORCOCONGATCDHVTE---CROPGYTGAFCEDLCPGKHGPOCEORC---PCONGV 245
DB 286 GSHHPCITNGTCTINAEPDQYRCTCPDQYSGRNCB-----KAEHACTSNPCANGGS 335
QY 246 CHHVTG--ECSRPSGMWMTVC-----GQP-----CPEGRFGKNC- 277
DB 336 CHEVPSGFECHPCPSMGSPICALIDECASNPCAAGTCVDQYDFECICPEQWVATCQ 395
QY 278 --SQEQC---CHNGTCDATG--QCHSPGYTGERCQ---DECPVGTGYLCAETCQCV 327
DB 396 LDANECEGRPLNAFSCKNLIGYCCDIPGMGINCHINVNDG-----RGQCQ 444
QY 328 NGCKCYH-VSG-ACICEAGFAGECEARLCEGLYGIKCDKRCRCHLENTHSCHPMSGEC 385
DB 445 HGGTCKDLVNGYOCYQCYGRGGRHCE-----LERDKC-ASSPCH-----SGG- 485
QY 386 ACKPMSGLYCNETCSPGYEACQ---QICS---CONGADCSVTGK--CTCAPGFKGI 437
DB 486 LCEDLDADGFHCH--CPQGFSGPLCEVDVLDCEBSPCRNGARCYNLEGDYCACPDDEGK 543
QY 438 DCST---PCPLGTGYTIGTSCRCCKKND-----VCSPIVDSCTCKAGHGVDCS 483
DB 544 NCSVPREPCPGA---CRVIDCGSDAGPMGTAAAGVCGP-HGRVCSQPG--NFS 595
QY 484 IRCPSGTWGPGCN-----LTCOLINGAC-NTLDG--TCTCAPGRGEKCE-----LP- 528
DB 596 CIDDSGTGYTCHENIDDCLOPCRNNGTCTIDEVDARCFRCPSGWEGELCDTNPNDCLPD 655
QY 529 -----CODGTYGLNCAER---CD---CSHADGCHPT--TGHCRLPG 562
DB 656 PCHSRGRCYDLVNDFYCACDDGMKGTCHSRERQCDAYTCSNGTCTYDSGDTFRACAPG 715
QY 563 WSGVHCDSVCAEGRMWGNCSLPCYCKNGASC--SPDDGICECAPGFRGTTCQRLCSPGY 620
DB 716 WKG---STCAVAK-NSSC-LPNPCVNGTGVSGASFSICIRDGWEGRTCT----- 761
QY 621 GHRCSQTPQCVAHSSGPHHITGL---CDLPGFTGALCN---EVCPSGRFGKNCAGIC 673
DB 762 -HNTNDCNPLPCYNGGIC--VDGVNMFRCBAGFAGPDCRINDECQS-----SPCAVGA 814
QY 674 TCTN--NGTCNPIRSCQCYPGWIGSDC-----SQPCPAHMGPNCIHTCNCHN 720
DB 815 TCYDEING-----YRCSPPGRAGPRCQEVIGFGRSCMRGTPFPH-GSSWEDCNS-- 865
QY 721 GAFCSAVDECKCTPWTGLYCTQRCPLGFGYGDALICQONGADC-DHISGQC---TC 776
DB 866 ---CRCLDGRDCKWVG--WKPCILA--GQPEALSACPLGQRLERAPGQCLRRPC 917
QY 777 RT-GFMGRHCEQKCP-----SGTYGCGROICDCLNNSTCDHI---TGTCYS-----PG 822
DB 918 EAMGEGGAEPSTPCLPRSGHLDNNCARLTLFNR---DHVPGTTVGAICSGIRSLPA 974
QY 823 WKGARCDQAGVIIVGNLNS-LSRTSTA-----LPADSYOIGALIGIILVLVLEL 873
DB 975 TRAVARDRLVLVLCDRASSASAVEAVSFPARDLPDSSLIGGAHAIVAAI----- 1027
QY 874 ALFTIIRHKOKKSSHPAVTYTPAMRVVNADYTISGTLPHSNGNANSHYFTNPSTYHTL 933

```

```

DB 1028 -----TORGNSSLIAVTEVKEVETVYTGSGSTGLLVPVLGA-----FSVYLMACY 1073
QY 934 TOCA--TSPVNNRDRMTYTK--SKNQ-----LFVNLKVNFGKR 970
DB 1074 VLYWMTTRKRKREKERSRLPRESANNQMAPLNFIRNPIRPGHKLVLQCKNFTPPR 1133
QY 971 -----GPVG 974
DB 1134 RADEALPGPAG 1144

```

Search completed: May 9, 2003, 11:59:20
Job time: 58.4565 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 9, 2003, 07:56:52 ; Search time 60.7648 Seconds

(without alignments)
2499.898 Million cell updates/sec

Title: US-10-092-390-2

Perfect score: 6744

Sequence: 1 MVTSLNSCLSFILLCWHT.....SSPKQDSGSSSSSSSSSE 1140

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
A.GeneSeq_101002:.*
1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3107.5	46.1	878	22	ABG08033 Novel human diagno
2	2506.5	37.2	1050	22	AA866267 Human TANGO 272 SE
3	1909	28.3	636	22	AA866269 Rat TANGO 272 SEQ
4	1522	22.6	384	22	AA875479 Human colon cancer
5	1466	21.7	321	22	ABG27639 Novel human diagno
6	1272	18.9	466	22	ABG22559 Novel human diagno
7	1252	18.6	434	22	AB866756 Drosophila melanog
8	1241.5	18.4	762	22	ABG08032 Novel human diagno
9	1192	17.7	474	22	AA72715 HFTU08 clone huma
10	1169	17.3	269	22	ABG08031 Novel human diagno

11	1034.5	15.3	2444	23	ABB07821	Constitutively act
12	1014.5	15.0	1473	23	AAE18208	Human NOTch protei
13	1012.5	15.0	2471	20	AA706816	Human NOTch2 (huma
14	1011	15.0	241	22	AA72712	HTLH44 clone huma
15	1002.5	14.9	587	22	ABG03826	Novel human diagno
16	980	14.5	2469	23	AAE18207	Human MOLA protei
17	978.5	14.5	2703	22	ABG20266	Drosophila melanog
18	974	14.4	2321	19	AAW49658	Human Notch3 prote
19	961.5	14.3	447	22	ABG66268	Human TANGO 272 S
20	953.5	14.1	1964	20	AAV95557	Mus musculus notch
21	940	13.9	1872	19	AAW68510	Partial human Notc
22	919	13.6	4618	22	AAW39043	Human polypeptide
23	895	13.3	209	22	ABG27638	Novel human diagno
24	813	12.1	833	21	AA85616	Murine nurse cell
25	813	12.1	833	22	AA860392	Murine nurse cell
26	808	12.0	865	23	AAW47668	Mouse nurse cell
27	808	12.0	866	22	AAW60393	MOLab protein sequ
28	808	12.0	866	22	AAW60394	Human nurse cell
29	802.5	11.9	831	23	ABB80609	Human nurse cell
30	800	11.9	871	22	AAW60395	Human sbg115B SRC
31	800	11.9	871	23	ABB80610	Human nurse cell
32	789	11.7	1238	23	ABB07823	Human notch agonis
33	788	11.7	1212	19	AAW44299	Human serrate 2
34	787	11.7	1055	19	AAW44298	Human serrate 2 pr
35	785	11.6	129	22	ABG22555	Novel human diagno
36	785	11.6	129	22	ABG27637	Novel human diagno
37	783.5	11.6	884	23	AAW47667	MOLab protein sequ
38	781	11.6	1257	17	AAW05834	Human Serrate-2 (H
39	779	11.6	1257	21	AAV59598	Skin cell protein,
40	779	11.6	296	22	AAW56069	Murine EGF family
41	779	11.6	296	23	ABW22629	Murine EGF family
42	779	11.6	299	21	AA76014	Murine EGF family
43	779	11.6	299	21	AA76017	Skin cell protein,
44	779	11.6	299	22	AAW59593	Skin cell protein,
45	779	11.6	299	22	AAW56016	Skin cell protein,

ALIGNMENTS

RESULT 1
ABG08033
ID ABG08033 standard; Protein: 878 AA.
AC ABG08033:
DT 13-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #8024.
XX
KW Human: chromosome mapping; gene mapping; gene therapy; forensic;
KW Food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YF:
PI WPI, 2001-639362/73.
DR N-PSDB; AAS72220.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations

PT responsible for genetic disorders or other traits and to assess
 XX biodiversity -
 PS Claim 20; SEQ ID No 38392; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polypeptide chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantifying a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

CC Sequence 878 AA:

Query Match 46.1%; Score 3107.5; DB 22; Length 878;
 Best Local Similarity 59.5%; Pred. No. 4.4e-155;
 Matches 594; Conservative 15; Mismatches 69; Indels 321; Gaps 13;

QY 338 ACICEAGFAGEGCEARALCEPEGLYGIKCDKRCPCHELEN--THSCHPMSEGCACKPGMSGLY 395
 DB 5 ALICQUTYA-----C-----ISAOLICPFAMEQOLVACHCHPMSEGCACKPGMSGLY 50
 QY 396 CHETSPGTYGACQOICSCQNGADGDSYTGKCTCAPGFKGIDCSTPCPLGYTGINCSSR 455
 DB 51 CHETSPGTYGACQOICSCQNGADGDSYTGKCTCAPGFKGIDCSTPCPLGYTGINCSSR 110
 QY 456 GCCKNDVAVSPYDGSCTCAGHGVDCSTRCPSGTMGFCNCINTCCLNGACNTLDGCT 515
 DB 111 GCCKNDVAVSPYDGSCTCAGHGVDCSTRCPSGTMGFCNCINTCCLNGACNTLDGCT 170
 QY 516 CAPGNRGEKCELPDQGTGTGLNCAERDCSHADGCHPTTGRCRLPGMSGVHCDVCAEG 575
 DB 171 CAPGNRGEKCELPDQGTGTGLNCAERDCSHADGCHPTTGRCRLPGMSGVHCDVCAEG 230
 QY 576 RRGPNCSLPCYCKNAGSCSPDDGICECAFEGTTCQRIKSCPGFYGHRCSTQCPCHSS 635
 DB 231 RRGPNCSLPCYCKNAGSCSPDDGICECAFEGTTCQRIKSCPGFYGHRCSTQCPCHSS 290
 QY 636 GPCHHITGCDCLPGFTGTLGNEVCPSSRGKNCAGICTNNGCNPIDRSCQCCPGMI 695
 DB 291 GPCHHITGCDCLPGFTGTLGNEVCPSSRGKNCAGICTNNGCNPIDRSCQCCPGMI 332
 QY 696 GSDSCQPCPP-----AHMGPNCIH-----TCCHNGAFCSAVDECKCTP-- 735
 DB 333 PAPTEPEYPLPLDLVSYFVGLAVTALNHNHLP.TGAGTASTRATATM-----ELSAAPIM 387
 QY 736 -----GWTGL-----YCTORCLPGFYGKDCALICQCONGADCDHISGQCTCRFG 779
 DB 388 GNVNALLAGQGSTALRRSPRHSCRAAASPWFYKDCALICQCONGADCDHISGQCTCRFG 447
 QY 780 FMGRHCEOK----- 788
 DB 448 FMGRHCEOKVRRPMDHRLMTLALGGGGVTTMRKTEFKSILFFWALPSSSYFWNVAOS 507
 QY 789 ----- 788
 DB 508 LKRSSAFPMALAEFGSHIGQYIKMGGLVAGOSLLPCAIVTVYSATIMIGMLSSSGT 567

QY 789 -----CPSGTGYGCCRQIC 802
 DB 568 ILGVQVSLNRNP.LKGLSSNACGLAVRDSLAPNSQGMKATFDEPSPLECPSPGYGCCRQIC 627
 QY 803 DCLNNSTCDHITGTCTCSPGKMGARCDAGVITIVGNLSLSTSTALPADSTQIQAIAIGI 862
 DB 628 DCLNNSTCDHITGTCTCSPGKMGARCDAGVITIVGNLSLSTSTALPADSTQIQAIAIGI 687
 QY 863 ILLVLYVFLALFFIIRHKQKRESSMPAYVTYTPAMRVNADYITSGTLPHSNGNANS 922
 DB 688 ILLVLYVFLALFFIIRHKQKRESSMPAYVTYTPAMRVNADYITSGTLPHSNGNANS 747
 QY 923 HFTNPSTHTLQCATSP-HVNNRDMRYTKSNQOLFVNLKNVNGKRGVDCDTGFLP 981
 DB 748 HFTNPSTHTLQCATSP-RSTGTGRMYVHEKK----- 781
 QY 982 ADMKRGVYLNELGALGDRSYGKSLKDLGKNSSEVNSSCSLSSSENPRATIKDPVLLP 1041
 DB 782 ----- 781
 QY 1042 KSECEGYEMKSPARDSPYAEINNSTSANRVYEVEPTVSYGVFSNNGRLSODPYDL 1101
 DB 782 QSTVC--ESMKSPARDSPYAEINNSTSANRVYEVEPTVSYGVFSNNGRLSODPYDL 839
 QY 1102 PRNSHIPCHYDLLPYRDSSSSPKQEDSGGSSSSSSSE 1140
 DB 840 PRNSHIPCHYDLLPYRDSSSSPKQEDSGGSSSSSSSSSE 878

RESULT 2

AA66267 standard; Protein; 1050 AA.

AA66267; (first entry)

Human TANGO 272 SEQ ID NO: 14.

Membrane associated protein; secreted protein; human; mouse; rat;
 INTERCEPT 340; TANGO 003; TANGO 347; TANGO 272; TANGO 295; TANGO 354;
 TANGO 378; skeletal disorder; cardiovascular disorder; renal disorder;
 haematopoietic disorder; neural disorder; hepatic disorder;
 neoplastic disease.

Homo sapiens.

WO200100673-A1.

04-JAN-2001.

29-JUN-2000; 2000WO-US18198.

30-JUN-1999; 99US-0345464.

(MILL-) MILLENNIUM PHARM INC.

Barnes TM, Fraser CC, Wrighton N, Myers P, Busfield SJ, Sharp JD;

WPT: 2001-050128/06.

N-PBDB; AAF27787.

Isolated secreted or transmembrane proteins are used for diagnosis and
 treatment of neoplastic and haematopoietic disorders e.g. T cell
 disorders, cancer and tumours -

Claim 9; Page 227-229; 294pp; English.

The present invention provides the protein and coding sequences for a
 number of membrane associated and secreted proteins from human, mouse
 and rat. The proteins are designated INTERCEPT 340, TANGO 003, TANGO 347,
 TANGO 272, TANGO 295, TANGO 254 and TANGO 378. The proteins are all
 involved in signal transduction and the sequences can be used in the

CC treatment of cardiovascular, renal, hepatic, neural, neoplastic, skeletal
CC and haematopoietic disorders.

XX Sequence 1050 AA;

Query Match 37.2%; Score 2506.5; DB 22; Length 1050;
Best Local Similarity 40.5%; Pred. No. 1.6e-123;
Matches 490; Conservative 111; Mismatches 345; Indels 263; Gaps 30;

```

QY 14 LLLCHWIGTASPLLEDPNVCSSHEISVYQESTYPRPFDIYTTCTDILNM----- 66
DB 9 LLLAVGLRAGTLPSPDPTSCFWESEFTTTTKEHSRPFSLPSEPC--RPWEGPHTCP 66
QY 67 -----FCTRHRYSTR-----TAY 80
DB 67 SPOTOKILLASRDSFCVAVGVOMRDRSALQPTGTNALSMPROPVYLASAPLASBPH 126
QY 81 RHGEKTYRRKQCCPGEFSEGEKCVPHADKCYHGRCAIAPNTCCQEPGNGTNCSSA-- 138
DB 127 TVVYKTDHROLRQCCCHGFSRGCYPLCAQECVHGRVAVANQCVGWMGRDDCCSAPN 186
QY 139 -----CDGDHMGPHCTSRCCCKNGALCNPTGACHCAAGFRGWRCEDECGTYGNDCHQR 194
DB 187 CLDPTGTYGPACQFRCQC--HGAPCDPQTGACFCFPAERTGSCDVSQSGT----- 237
QY 195 CQCQNGATCDHVTGECNCPPTGTGAFCEDLCPKGHGRCQRCPCQNGCVCHVTEGCS 254
DB 238 -----SSGFFC-----PSTH-----PCQNGVGFQTPQSGCS 262
QY 255 CPSCMGTVCGQPCPEGRFRKNCQEQCHNGGTCDAATGQCCHGSPGYTGERCODECPVG 314
DB 263 CPPEMNGTICSLPCEPHGHPNCQEQCHNGGTCDAATGQCCHGSPGYTGERCODECPVG 322
QY 315 TYGVLCATCOCVNGKCYHVSACLDGAFGERCEARLCEBLYGKICDKRCPCHLEN 374
DB 323 RFGDCAETCCADARCFRANGACLEHGFETGDRCTDLCPDFFYGLISCAQPCCTCREH 382
QY 375 THSCHPMSGECAKPGMSGLTYGSGPFGFAGACQOICSCQNGADDDSVYKCTCARGF 434
DB 383 SLSCHPANGECSCLPNGAGLCHNCSCPDOTHGRCQCHCLDHLHGVCATISGLCQACAGY 442
QY 435 KGIDCSTPCPLGTGYNCSRCGKNDAYCSPVDSCTCKAGWGVDCSIRCPSTGWF 494
DB 443 TGPHCASLCPDPTGVNCSARCSENAIACSPIDGECCKGMRGNSVYCPGTWFS 502
QY 495 CNLTCCQCLNGACNTLDGTCTCARGMEKCELPQODGTGYNCAERDCSHADCHPTT 554
DB 503 CNAQCQCAHEAVYCSPTGACTCTPGWHGANCOLPCPKQGFEGCASRDCDHSDCDPVH 562
QY 555 GHCRCLPGMSGVHDDVCABEGRMNCSLPYCKNGASGSPDDGICCECAPGFRGTORI 614
DB 563 GRCCQAGMGMARGHLSCPBELMGVNCNNTCTCKNGGCLDPENGCAPAFRPSGORS 622
QY 615 CSPFGYHGRSQTCRQCVHSSGPNHITGLDCLPFGTALCNEVCSGFRGKNCAGICT 674
DB 623 CQPRYGRK-----CYP-----CK 636
QY 675 CTNNGTCNPIDRSCQCTPGWIGSDCSQPCRAHWPNCITTCNCHNGAFCSAYGECKCT 734
DB 637 CANHSFCHPNSMTCTYCLAGMTGPRDSCQPCRGHNGENACQCCCHNGGTCHPDGSCICP 696
QY 735 PGWGTGLCTONCRPLGFTGKDCALICQONGADCHISGQCTCRGFMNRHCEQKCPSTY 794
DB 697 LGWTHHNLCECPGTGANCSPQCGPGEKC-----HPE----- 732
QY 795 GYGCROICDCLNNTCHITGTCTCSPGMKARGCDQAGVILVGNLNSLSRSTALPAD-- 852
DB 733 -----TGACVCPGHSAPCR-----IG-----IQEPFTYMPPTPV 763
QY 853 SY-OIGAIAGIILVVLVFLALFIYRHKQKESMPAVYTTPAMRVVADYTSIGT 911
DB 764 AYSIGAVNTGIVAGSLVALVAFIYGRHMQKKEHHNLHVAVSSG-RLDGSEVYMDV 822

```

```

QY 912 LPHSNGNANSHYFTNPNSTYHTLQCATSPHYNNRDMRTVTKSKNOLFEVNLKN-VNPKGR 970
DB 823 PP-----SYSHYSNPSYHTLSQCSPPNPPNK-----VGPPLFASLQNERG-- 866
QY 971 GPVG-DCTGTLPADMKH-----GGYLNELAFGLDRSYMKSL-----KDLGNSEY 1016
DB 867 GAQCHDHNHTLPLADMKHRRPEPPPLDR-GSSRLDRSYSYSGNPGPFYDKGLISEEL 925
QY 1017 NSSMCSLSSSNPYATTKDPPVLLPKSSECIYEMKSPARDSPYAEIINSTSANRNYE 1076
DB 926 GASVASTL-SSSNPYATTKDPPVLLPKSSECIYEMKSPARDSPYAEIINSTSANRNYE 981
QY 1077 VEPTVSYVQGVFSNNGRL-----SODP-----YDLPKNSHIPCHYDLIPYVDS 1119
DB 982 -QOPQDRDSCTIEQPSPLIHDRDSVGSOPPLPPGLPCHGHTSPKNSHILPGHYDLPVYRHP 1040
QY 1120 SSSP-KQED 1127
DB 1041 PSPPLRRQD 1049

```

RESULT 3

AAB66269
ID AAB66269 standard; Protein: 636 AA.

AC AAB66269;

DT 05-APR-2001 (first entry)

XX Rat TANGO 272 SEQ ID NO: 20.

DE Membrane associated protein; secreted protein; human; mouse; rat;
KW INTERCEPT 340; MANGO 003; MANGO 347; TANGO 272; TANGO 295; TANGO 354;
KW TANGO 378; skeletal disorder; cardiovascular disorder; renal disorder;
KW haematopoietic disorder; neural disorder; hepatic disorder;
KW neoplastic disease.

OS Rattus sp.

PN W0200100673-A1.

PD 04-JAN-2001.

PF 29-JUN-2000; 2000WO-0518198.

PR 30-JUN-1999; 9905-0345464.

PA (MILL-) MILLENNIUM PHARM INC.

PI Barnes TM, Fraser CC, Wrighton N, Myers P, Busfield SJ, Sharp JD;

DR WPI; 2001-050128/06.

DR N-PSDB; AAF27791.

PT Isolated secreted or transmembrane proteins are used for diagnosis and

PT treatment of neoplastic and haematopoietic disorders e.g. T cell

PT disorders, cancer and tumours -

PS Claim 9; Page 238-240; 294pp; English.

CC The present invention provides the protein and coding sequences for a
CC number of membrane associated and secreted proteins from human, mouse and
CC rat. The proteins are designated INTERCEPT 340, MANGO 003, MANGO 347,
CC TANGO 272, TANGO 295, TANGO 254 and TANGO 378. The proteins are all
CC involved in signal transduction and the sequences can be used in the
CC treatment of cardiovascular, renal, hepatic, neural, neoplastic, skeletal
CC and haematopoietic disorders.

XX Sequence 636 AA;

Query Match 28.3%; Score 1909; DB 22; Length 636;
Best Local Similarity 45.1%; Pred. No. 2e-92;
Matches 328; Conservative 77; Mismatches 212; Indels 110; Gaps 9;


```
XX 18-FEB-2002 (first entry)
DT
XX
DE Novel human diagnostic protein #27630.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI: 2001-639362/73.
XX
DR N-PSDB; AAS91826.
XX
PS Claim 20; SEQ ID No 57998; 103bp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SO Sequence 321 AA;

Query Match 21.7%; Score 1466; DB 22; Length 321;
Best Local Similarity 90.3%; Pred. No. 1.7e-69;
Matches 241; Conservative 0; Mismatches 0; Indels 26; Gaps 2;

OY 137 SACCDDHNGPCHTSCRCCKNGALCNPIGACHCAGFGMRCEDECGGTGNDCHQOQ 196
DB 9 SACCDDHNGPCHTSCRCCKNGALCNPIGACHCAGFGMRCEDECGGTGNDCHQOQ 68
OY 197 CONGATCDHYTGECRCPCGYTGAFCEDELCPGKHGPOCEORPCONGVCVCHHTGECSCP 256
DB 69 CONGATCDHYTGECRCPCGYTGAFCEDELCPGKHGPOCEORPCONGVCVCHHTGECSCP 128
OY 257 SGM-----GTVCGQPCPEGRFGKNCSCQCGCHNGTCDATGQCHC 298
DB 129 SGMMLSPFGWPIKXSKSLXMGITVCGQPCPEGRFGKNCSCQCGCHNGTCDATGQCHC 188
```

```
OY 299 SPGYTGER-----CODECPVGYGVLCALTCQCVNGKCYHVSACLCEAGPAGRC 350
DB 189 SPGYTGERAAVPDVRRKCODECPVGYGVLCALTCQCVNGKCYHVSACLCEAGPAGRC 248
OY 351 EARLCEGLYGIKCDKRCPCHELENTHS 377
DB 249 EARLCEGLYGIKCDKRCPCHELENTHS 275

RESULT 6
ABG22559
ID ABG22559 standard; Protein: 466 AA.
XX
AC ABG22559;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #22550.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI: 2001-639362/73.
XX
DR N-PSDB; AAS86746.
XX
PS Claim 20; SEQ ID No 52918; 103bp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SO Sequence 466 AA;

Query Match 18.9%; Score 1272; DB 22; Length 466;
Best Local Similarity 76.9%; Pred. No. 3.5e-59;
```


CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations in
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 762 AA:

Query Match 18.4%; Score 1241.5; DB 22; Length 762;
 Best Local Similarity 44.0%; Pred. No. 2,2e-57;
 Matches 267; Conservative 16; Mismatches 81; Indels 243; Gaps 17;

QY 545 SHADGCHPTGHCRCILPGNSGVHCDYVCAEGRWGPNCSLPYCKNAGSCSPDDGICECAP 604
 DB 14 SHSVISGAAAGEC-----GGVHCDYVCAEGRWGPNCSLPYCKNAGSCSPDDGICECAP 67
 QY 605 GRRGTTTCORICSPGFYGHRCISQTCPCVHSSGRCRHITGLCDCLGFTALCNEV----C 660
 DB 68 GRRGTTTCORICSPGFYGHRCISQTCPCVHSSGRCRHITGLCDCLGFTALCNEVSSQC 127
 QY 661 PSGRGKNCAGICTCTNNCTGNCNPNIDRSCOCYPGWIGSDSCQPCPAHMPNCIH-TCNCH 719
 DB 128 PSGRGKNCAGICTCTNNCTGNCNPNIDRSCOCYPGWIGSDSCQPCPAHMPNCIH-TCNCH 183
 QY 720 NCAFCSAIVGECKCTPGWTGLYCTGRCPLGFTYGKDCALICQONGADCDHISGQCTCRPG 779
 DB 184 N-----TCOCEPGW-----GGTNCSSVKKQSP----- 205
 QY 780 FNGRHCEOKCPSGTYGYGRCOLCDLNNSTCHITGTCTCSPGWKGARGDQAGVILVGNL 839
 DB 206 -----VCESKK-----CEP-WERRPC-----GGI 223
 QY 840 NSLSRTSTALPADSYOIGALIGIILVVLFLALFIYRHKOKGESSMPAVTYTPAM 899
 DB 224 H----- 224
 QY 900 RYVADYITSGTLPHSNGNANSHYFTNPSTHTLQCATSPHYNNDRMTVTKSKNNQLF 959
 DB 225 -----WDIAGX----- 230
 QY 960 VNLKVNPGKRGVGDCTGTLPADMKHGGYLNEIGAFGDRSYMGSKLKDIGNSEYNS 1019
 DB 231 -----PQ-----KHGGYLNEIGAFGDRSYMGSKLKDIGNSEYNS 267
 QY 1020 NCSLSSENRYATIKDPVLPKRSB---CG---YVEMKSPARRDSPYAEINNSTANRN 1073
 DB 268 NCSLSSENRYATIKDPVLPKRSB---CG---YVEMKSPARRDSPYAEINNSTANRN 1073
 QY 1074 VTEVEPYVSVGVFNNRGLSODPYDLFKNSHIPCHYDLFVRDSSSPKQEDSGGSSS 1133
 DB 328 YSECCPR-SIOQXMASFPESIXP-----PKNSHIPCHYDLFVRDSSSPKQEDSGGSSS 381
 QY 1134 NSSSSSE 1140
 DB 382 NSSSSSE 388

RESULT 9

AAV72715
 ID AAV72715 standard: Protein; 474 AA.
 XX
 AC AAV72715;
 XX
 DT 31-MAY-2001 (first entry)
 XX
 DE HFCU08 clone human attractin-like protein.
 XX
 KW Human; attractin-like protein; HFCU08 clone; food additive;
 KW food preservative; gene therapy; cancer; neuroprotective; cytostatic;
 KW immune disorder; Addison's disease; allergy; cerebroprotective;
 KW autoimmune haemolytic anaemia; autoimmune thyroiditis; noctropic;
 KW diabetes mellitus; Crohn's disease; multiple sclerosis; cardiac;
 KW rheumatoid arthritis; ulcerative colitis; cardiovascular disorder;
 KW myocardial ischaemia; congestive heart failure; pericarditis; viral;
 KW wound healing; neurological disease; cerebral anoxia; antibacterial;
 KW Alzheimer's disease; communicative disorder; epilepsy; ophthalmological;
 KW infectious disease; immunosuppressive; antiproliferative; vasotropic;
 KW fungicide; chromosome 1p36.3.
 XX
 OS Homo sapiens.
 XX
 FH Key
 FH Location/Qualifiers
 FH Region 14..20
 FT /note= "Immunogenic epitope"
 FT 37..52
 FT /label= EGF-like_domain
 FT 55..66
 FT /note= "Immunogenic epitope"
 FT 80..95
 FT /label= EGF-like_domain
 FT 86..92
 FT /note= "Immunogenic epitope"
 FT 123..138
 FT /label= EGF-like_domain
 FT 152..160
 FT /note= "Immunogenic epitope"
 FT 167..182
 FT /label= EGF-like_domain
 FT 176..181
 FT /note= "Immunogenic epitope"
 FT 210..225
 FT /label= EGF-like_domain
 FT 262..273
 FT /note= "Immunogenic epitope"
 FT 279..287
 FT /note= "Immunogenic epitope"
 FT 296..311
 FT /label= EGF-like_domain
 FT 325..332
 FT /note= "Immunogenic epitope"
 FT 339..354
 FT /label= EGF-like_domain
 FT 382..397
 FT /label= EGF-like_domain
 FT 392..397
 FT /note= "Immunogenic epitope"
 FT 404..414
 FT /note= "Immunogenic epitope"
 FT 429..434
 FT /note= "Immunogenic epitope"
 FT 461..474
 FT /note= "Immunogenic epitope"
 XX
 PD WO200116156-A1.
 XX
 PD 08-MAR-2001.
 XX
 PF 29-AUG-2000; 2000WO-US23663.
 PF
 XX 30-AUG-1999; 99US-0151348.
 PR
 XX

Qy 456 CGCKNDVAVSPVDSCTCTKAGHGVDCSIRCPSTGFGCNLTCCOLNGACNTLDGCT 515
 Db 111 CGCKNDVAVSPVDSCTCTKAGHGVDCSIRCPSTGFGCNLTCCOLNGACNTLDGCT 170
 Qy 516 CAPGRGKCELPCCDGTGYGLNCAERCDSCSHADGCHPTTGCHRCCLPGMS 564
 Db 171 CAPGRGKCELPCCDGTGYGLNCAERCDSCSHADGCHPTTGCHRCCLPGMS 219

RESULT 11
 ABB07821
 ID ABB07821 standard; Protein: 2444 AA.
 AC ABB07821;
 DT 03-JUL-2002 (first entry)
 XX Constitutively active notch-1 protein.
 DE Cell differentiation; notch; epidermis; cytosolic; dermatological;
 KW epithelial; skin; cancer; gamma secretase; notch-1.
 XX Synthetic.
 OS Key Location/Qualifiers
 FH MISC-difference 891 /label- Gly or Ala
 FT /note- "encoded by GSS"
 FT MISC-difference 1763 /label- Glu, Arg, Pro or Leu
 FT /note- "encoded by CNG"
 FT MISC-difference 1787 /label- Thr, Ala, Pro or Ser
 FT /note- "encoded by NCC"
 XX WO200218544-A2.
 XX 07-MAR-2002.
 XX 31-AUG-2001; 2001WO-US27246.
 XX 31-AUG-2000; 2000US-229614P.
 XX (LOYO) UNIV LOYOLA CHICAGO.
 XX Nickoloff BJ, Miele L;
 DR WPI: 2002-339659/37.
 DR N-PSDB: ABLA0768.
 XX Inducing differentiation of epithelial cell useful for inducing barrier
 PT formation within epithelium for treating psoriasis, sunburn, involves
 PT exogenously providing a source of a Notch agonist to the epithelial
 PT cell.
 PS Claim 10; Page 50-66; 101pp; English.

The invention relates to a method of inducing differentiation of
 at least one epithelial cell. The method involves exogenously providing
 at least one source of at least one Notch agonist to at least one
 epithelial cell, whereby the Notch pathway is activated within at least
 one epithelial cell so that the differentiation of the cell is induced.
 Methods of producing differentiated epidermis: for assaying for genetic
 propensity of a patient to develop a disorder associated with epithelial
 barrier formation; for retarding progression of skin cancer and for
 diagnosing aggressive melanoma are also provided. The methods are useful
 for inducing differentiation of at least one epithelial cell e.g. a
 keratinocyte or a pre-malignant cell, in vivo or ex vivo. The method is
 useful for inducing differentiation of epithelial cell within cutaneous
 epithelium such as oral mucosal epithelial tissue, cornea epithelial
 tissue, gastrointestinal epithelia, urogenital epithelia, or respiratory
 epithelia. The methods are useful retarding the progression of skin

CC cancer such as aggressive melanoma, aggressive cutaneous T-cell lymphoma
 CC (CTCL), aggressive squamous cell carcinoma, or aggressive basal cell
 CC carcinoma, by preferably administering an antagonist of the Notch
 CC pathway such as gamma secretase inhibitor. The present sequence
 CC represents a constitutively active notch-1 protein.
 XX
 SO Sequence 2444 AA;

Query Match 15.3%; Score 1034.5; DB 23; Length 2444;
 Best Local Similarity 25.8%; Pred. No. 4.8e-46;
 Matches 316; Conservative 83; Mismatches 304; Indels 523; Gaps 73;

Qy 94 CCGPFGSEEMCVPHCADKCYHRC-----IANTCCCGRGSGMTSSACDDH 143
 Db 89 CALGF--SGPLCLPPLDNACLNPCRNCGTCDLTLTTEYKRCRPGMSGSKCOA----- 141
 Qy 144 WPHCTSRQCKNGALCNPTGA--CHCAAGFRGRC--DREBQ--TYGNDHQ- 193
 Db 142 --DPCASN-PCANGGQCLPFEASYICHCPSPHPTCRQDYNEGQKPRLCRHGGTCHNE 198
 Qy 194 ----RC-----OQNGATC--DHYTGRCRCPPTTGAFCE--- 222
 Db 199 VGSYRCVCRATHHTGPNCEPRYPVPCSPONGGTCRPTGDTVHECACLPFTGQNCENI 258
 Qy 223 DLCPPG--KHGPRC-----EQRCP-----CONGVCHHYTG- 251
 Db 259 DDCPRNNCKNGACVGYVNTATNCPRPMTGQYCTEDVDECOLMNAQNGTCHNTHG 318
 Qy 252 -ECSPSGMWTVCQ-----PCPRGRFKNC--SQEC-- 281
 Db 319 YNCVAVNGMTGEDCSENIDDCASACFAGATCHDRVASFYCECPHGRGLCHLINDACTS 378
 Qy 282 -QCHNGTCD--ATGQ--CHCSPTGTBERQ--DEPVTGYTLCATFOQVNGKCY 333
 Db 379 NPCNENSGNDNPVNGKATCTCPSGTYGPACSDVDECSLAN-----PCBAGKCI 430
 Qy 334 HVSGA--CLCEAGFAGECEARLCEGLYGIKIDRC--PCHLEHTSCHPMSE--CA 386
 Db 431 NTLGSFECQCLDGYTPRCEIDV-----NECSNRC--QNDATCLDQIGERPCM 477
 Qy 387 CKPGNSGLYC-----NE--TCSPGFYGEACQ--QICS--C 415
 Db 478 CMPEYEGVHCENVTEDECASSPCLHNGRCLDKINEPQCEPFGTGLCOYDVEDCASPNC 537
 Qy 416 ONGADC-----DSV--TGKCTCAPRGK 436
 Db 538 KNGAKCLDGPNTYTCVCTEGYTGHCVEYDIDCEDDPCHYSGCKNGVATFTCLCRPYTG 597
 Qy 437 IDCSF-----PCPL-----GTGINS-----SRGCKNDVACS 465
 Db 538 HHCETNINECSSQPCRLNGTCCDPNATVLCRLKTTGPNCENINDDCASSPCDSGTCLD 657
 Qy 466 PVDG-SCYCKAGHGVDCSIR-----CPSGTWFGCNL-----TC----- 499
 Db 658 KIDGECACEPGEYTSMSNSNIDECAGNPNCHNGTCEGIDINGFTCRCEGYHDFCLSEV 717
 Qy 500 -----OCLNGACNTLDG-TCTCAPGRGKCELP----- 527
 Db 718 NECSNPNPVHACRSLNGRYKCDDPGWSGNTCLNNNECESNPNVNGTCKDMTSGIVC 777
 Qy 528 PCODTYVLNCAERD-----CSHADG-HPYTGH--CRCLPGMSGVCHDSV---CARG- 575
 Db 778 TCRBFSFSPNCTNINECASPNCLNKGTICIDVAGYKKNCLLPYTAGCEVYLAAPCAASP 837
 Qy 576 -RMGPNC-----SLPYC--KNGASCSPDDGICECAGFRGTTCORI-----CS 616
 Db 838 CRNGECRQSDYDESFSCVCPAGAKGTCEVDINECVLSPCRHAGASCONTHGYRCHQ 897
 Qy 617 PGFTGHRCSQCPQCVHSSGRCH-----TGLCCLLPGFTGALCNE----- 658
 Db 898 AGYSGRNCETDIDC--RPNCHNGSGCTDGINATFACCLGPFRTFCEEDINECASDPC 955
 Qy 659 -----VPSGRFRKNCAG--ICT--CTNNGTCNPIDR-----SCQCYPG 693

```

Db 956 RNCAGCTCDVSYCTCPAGFSGIHCENNTPDCTESSCFNGATC--VDGINSFTCLCPG 1013
QY 694 WIGSDC-----SOP-----CPRAHNGPNC---IHCNC---CHNGA 722
Db 1014 FTGSCYCHVYNECDSPCLLGGTCQDGRGLHRTCTPGGYTGPNCLVHWCDDSPCKNG 1073
QY 723 FC-----SAYDECKCTPGMTGLYCTQ----- 744
Db 1074 KCMQTHQY--RCECPSGMTGLYCDVPSVSCVAAHQSGVAVRLCHGGLCVDAAGTTH 1131
QY 745 -RCPLEGYGCDAIY---CO---CQNGACD-DHISG--OCTCRGFMGRHCEOK----- 788
Db 1132 CRCQAGYTGVCYCDLVDEGSPSPCQNGATCTDYLGYSCKCVAGHGVNCSSEEDCLSH 1191
QY 789 -----CPSGTYGCGROI CD-----CLNNSTC-DHITG 815
Db 1192 PCQNGCTCLDLPNTYKSCSRGTGVCCEINVDCCNPVDPVSRSPKCFNNGTCVDVGG 1251
QY 816 -TCYSPGMKARGCDQAGVIIVGNLN 840
Db 1252 YSCTCPGPFYGERCE-----GPDVN 1270

RESULT 12
AAE18208
ID AAE18208 standard; Protein; 1473 AA.
XX
AC AAE18208:
XX
DT 07-MAY-2002 (first entry)
XX
DE Human MOL1b protein.
XX
KM Secreted molecule; MOL1b protein; MO1x: cardiomyopathy; atherosclerosis;
KM diabetes; chromosomal disorder; albinism; anemia; psoriasis; scarring;
KM liver cirrhosis; neurodegeneration; osteoarthritis; organ rejection;
KM cerebral thrombosis; hypertension; systemic lupus erythematosus; asthma;
KM immune disease; leishmaniasis; immunodeficiency; rheumatoid arthritis; ulcer;
KM HIV; human immunodeficiency virus; hepatitis; haemostatic disease; pain;
KM hematopoietic; thrombolytic; thrombocytopenia; Alzheimer's disease;
KM Parkinson's disease; Huntington's disease; muscular disease; stress;
KM ocular disease; growth disorder; depression; epilepsy; contraceptive;
KM vulnerability; osteopathy; haemostatic; tranquiliser; antidepressant;
KM analgesic; vasodilator; hypotensive; gene therapy; chromosome 1.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1..25
FT 26..1473
FT Protein /label= Mature_MOL1b_protein

MO200206339-A2.
XX
PD 24-JAN-2002.
XX
PF 03-JUL-2001: 2001MO-US21249.
XX
PR 03-JUL-2000: 2000US-215854P.
PR 03-JUL-2000: 2000US-215856P.
PR 03-JUL-2000: 2000US-215902P.
PR 07-JUL-2000: 2000US-216585P.
PR 07-JUL-2000: 2000US-216586P.
PR 07-JUL-2000: 2000US-216732P.
PR 17-JUL-2000: 2000US-218632P.
PR 17-JUL-2000: 2000US-218932P.
PR 27-JUL-2000: 2000US-221285P.
PR 14-FEB-2001: 2001US-268734P.
XX
PA (CURA-) CURAGEN CORP.
XX

```

```

PI Spaderna SK, Tchernev V, Liu X, Shenoy S, Spytek K, Zerhusen B;
PI Pattirajan M, Tappier RJ, Rastelli L, Grosse WM, Ssekere ES;
PI Alsobrook J, Lepley DM, Shen L, Burgess CE, Shinkels RA;
PI Padigaru M;
DR WPI: 2002-155038/20.
DR N-PSDB; AAD28942.
XX
PT Nucleic acids encoding secreted polypeptides, designated MOLX
PT polypeptides, useful for treating a MOLX-associated disorder, e.g.
PT cardiomyopathy, atherosclerosis, diabetes and metabolic disorders -
XX
XX Claim 1; Page 15-16; 223pp; English.
XX
CC The patent discloses nucleic acid sequences encoding novel secreted
CC molecule (MOL) polypeptides, designated MOLX polypeptides (i.e. a MOL
CC protein where X is an integer from 1 to 8). Sequences of the invention
CC are useful for treating or preventing a MOLX-associated disorder in
CC humans. They are useful for treating or preventing cardiomyopathy,
CC atherosclerosis and disorders related to cell signal processing and
CC metabolic pathway modulation. The MOLX antibodies are useful for
CC treating or preventing diabetes and disorders related to cell signal
CC processing and metabolic pathway modulation. MOLX sequences are useful
CC for the treatment or diagnosis of other MOLX-associated disorders, e.g.
CC chromosomal disorders, albinism, anaemia, liver cirrhosis, psoriasis,
CC scarring, neurodegeneration, osteoarthritis, organ rejection, cerebral
CC thrombosis, ischaemia, hypertension, systemic lupus erythematosus, immune
CC diseases, immunodeficiency, HIV (human immunodeficiency virus), viral,
CC bacterial and fungal infections, hepatitis, rheumatoid arthritis, asthma,
CC haematopoietic, thrombolytic, haemostatic diseases, thrombocytopenia,
CC ulcers, Alzheimer's disease, Parkinson's disease, Huntington's disease,
CC ocular disease, muscular diseases, growth disorders, loss of libido,
CC stress, depression, pain and epilepsy. They are useful for preventing
CC chemotherapy side effects and as contraceptives. Sequences of the
CC invention are also useful for gene therapy. The present sequence is
CC human Notch-like protein, MOL1b. MOL1b gene is localised on chromosome 1.
XX
SQ Sequence 1473 AA:
XX
Query Match 15.0%; Score 1014.5; DB 23; Length 1473;
Best local similarity 23.5%; Pred. No. 3,3e-45;
Matches 339; Conservative 102; Mismatches 361; Indels 643; Gaps 72;
QY 93 QCCPGEYSEGMCVPHCADKCVHGR-CIADPNTQ-----CEPGMGTCSSACDG 141
Db 91 RCASGF--TGEDCQYSTSHPCFVSRLPCLNGCTGMLSRDYECTCGYGTGKCE----- 142
QY 142 DHNGPHCTSRCCCKNGALCNPITG--ACHCAAGRGMRCDRECGTYGNDCHQROCCN 139
Db 143 -QMTDACLH-PCANGSTCTTVANQFSCKCLTGFTGCKE-----TDVNECDIPGHCH 194
QY 200 GATCDHYTG--EGRCPGYTGARCELD-----CPPGKHG 231
Db 195 GGTCTLNPFSYGCOCPPGFTGYCDSLYPCARSPCVNGSTCTNOTGDTFTECNCLPFE 254
QY 232 PQCEQ-----RCP-----CONGV 245
Db 255 STCERNIDDCPNHRCQNGVGVVNTYNCRCPPMTGQPCFTDVEDCLIPNACONGGT 314
QY 246 CHHYTG--ECSRPSGMMGTVCQ-----PPBGRFGNRC- 277
Db 315 CANRNGGYGCVVNGWSDGDCSENIDDCAPASTPGSTCIDRYASFSQMPBCKAGLCH 374
QY 278 -SQEC---OCHNGGTCD--ATGO--CHCSPGYTGERCQ---DECPYGTGVLCIAETCQC 336
Db 375 LDDACISNPKHAKALCDTNLNLNGOYICTCPQGTAKGADCTEDVDECA--ANSNPC 427
QY 327 VNGKCYHSGA--CICEAGFAGECE-----ARLCEPGLY 360
Db 428 EHAGKCVNTDGAHFHCECLKGYAGRCMDINECHSDPCONDATCLDKIGFTCLCMGFK 487
QY 361 GIKCDKR-----CP-----CHLE-----NTHSC 378

```


Db 65 CTCANGACSPIDGSCCTPGWLGDITCLPCPDGTGFLNCSEHDCSHADGCDPVTGHC 124

Query Match	14.98;	Score 1002.5;	DB 22;	Length 587
Best Local Similarity	31.08;	Pred. No. 5.7e-45;		

Matches 214; Conservative 49; Mismatches 178; Indels 249; Gaps 19;

```

OY 31 PIVCSHMEISYTVQESYHPFDQIYYTSTDLNFKTRHRVSRTRYRHEKTMRR 90
Db 5 PTVLCIMAOAYVQ-----RNNHIFWNEKEHG-----HG 34
OY 91 KSQCCPGFEYSEGMCVPHCADKCVHGRCIAPNTCOEPMGNTCS----- 136
Db 35 KSGSC-----HNGASC-----SADGACHCTPGWTLGCTQRKPHLASOPLRI 78
OY 137 -----SA 138
Db 79 PCCGLATVGIQTSREGMQAPGLVVPDSCPTREELCRGSSRPDMIQIGIDKPKVLQ 138
OY 139 CDGDHMGPHCTSRCCCKNACLPITGACHCAAGFRGRCEDRCQGYGANDCHORCQ 198
Db 139 CPAAFEGKDCGRVCCQNGASCDHISGKTCRTFTGQHCEQRCAPGTFGYGCCQDLCECM 198
OY 199 NGATCDHYTGECRPPGYTGAFCE-----LCPPG-----KHQPO----- 233
Db 199 NNSTCDHYTGTCYCSFGKGTGRCDOGIMLLFLYCAAGPICLASAAEREGPRGSPCL 258
OY 234 ---CEORCP-----CONGVCHVTEGECSCPSGMMGTVCQ 266
Db 239 LHTCHERRAPATTPSQDLTDHYLRFESMPIMVLTCLQGA---FPGSPGRPGXTMAPLCGM 314
OY 267 -----PCEGRFRGKNSQEQCHNGGTCDAATGQCHCSPGYTERCODECPVGT 316
Db 315 NVNRRGTHELGDSDHMGPHCSNRCCQNGALCNPTGACVCAAGFRGRCCELCAPGTH 374
OY 317 GVLCAETCQCVNGKCYHVSAGCLCEAGFAGERCEARLCPBGLYGIKCDKRCPLHENTH 376
Db 375 GKGCQLPCQCHRGASCDPRAGECLAPGYT-----GY----- 407
OY 377 SCHPUSGECACKPGHSGLYCNETSPGFYGAQOICSCQNGADCDVYTKCTCAPFGK 436
Db 408 -CHPYTGACTCOPGWSGHHCNESCVPVGYGDCQLPCTCQNGADCHSITGGCTCAPFGM 466
OY 437 IDGSPCLPGLTYGINCSSRCGCKNDAYCSPVDSCTCKAGNHGVDCTRCPSGTWGRGN 496
Db 467 EVCAYSCAAGTYGPNCSICSCNNGGTCTPIDGSCCTKEG--NVP-SLPSPLY----- 518
OY 497 LTCCLNGACNTLDGTCTCAPGWRGEKCELPQDGTGYNLCARCDSCSHADGCHPTGH 556
Db 519 -----EHIPQVYLPABG-----SODGTGFLNCSEHCDCHADGCDPVTGH 558
OY 557 CRCLPGNSGVCHDSVCAEG---RWGPNCSL 583
Db 559 CCCLAGWTDIQ-----EGFLEKEGPKRTL 582

```

Search completed: May 9, 2003, 11:54:35
 Job time : 80.7648 secs

Fri May 9 13:01:41 2003

us-10-092-390-4.rai

Page 1

GenCore version 5.1.4 p5 4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 9, 2003, 11:53:35 ; Search time 11.5435 Seconds

(without alignments)
1493.646 Million cell updates/sec

Title: US-10-092-390-4
Perfect score: 3601
Sequence: I WYSLNSCSFICLLCHMT.....HDSVCAEGRMGNCLPCY 586

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: Issued Patents NA:
1: /cgn2_6/prodata/1/1aa/5A.COMB.pep.*
2: /cgn2_6/prodata/1/1aa/5A.COMB.pep.*
3: /cgn2_6/prodata/1/1aa/5A.COMB.pep.*
4: /cgn2_6/prodata/1/1aa/5A.COMB.pep.*
5: /cgn2_6/prodata/1/1aa/5A.COMB.pep.*
6: /cgn2_6/prodata/1/1aa/5A.COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	779	21.6	239	US-09-188-930-332	Sequence 332, App
2	759	21.1	239	US-09-188-930-192	Sequence 192, App
3	719	20.0	2523	US-08-185-432-18	Sequence 18, App1
4	719	20.0	2523	US-08-899-232-3	Sequence 3, App1
5	678	18.8	1193	US-08-400-159-10	Sequence 10, App1
6	678	18.8	1193	US-08-611-729A-10	Sequence 10, App1
7	676	18.8	2703	US-08-185-432-19	Sequence 19, App1
8	676	18.8	2703	US-08-899-232-4	Sequence 4, App1
9	666.5	18.5	2471	US-08-185-432-16	Sequence 16, App1
10	666.5	18.5	2471	US-08-083-590A-19	Sequence 19, App1
11	666.5	18.5	2471	US-08-532-384-19	Sequence 19, App1
12	666.5	18.5	2471	US-08-899-232-1	Sequence 1, App1
13	663	18.4	2556	US-08-185-432-17	Sequence 17, App1
14	663	18.4	2556	US-08-899-232-2	Sequence 2, App1
15	662	18.4	2556	US-08-083-590A-20	Sequence 20, App1
16	662	18.4	2556	US-08-532-384-20	Sequence 20, App1
17	659	18.3	1864	US-09-467-997-1	Sequence 1, App1
18	636	17.7	1810	PCT-US95-11684-4	Sequence 4, App1
19	635	17.6	2199	PCT-US95-11684-2	Sequence 2, App1
20	634.5	17.6	1148	US-08-882-046-4	Sequence 4, App1
21	634	17.6	1055	US-09-214-278-2	Sequence 2, App1
22	633	17.6	1065	US-08-400-159-8	Sequence 8, App1
23	633	17.6	1212	US-09-214-278-3	Sequence 3, App1
24	633	17.6	1257	US-08-611-729A-8	Sequence 8, App1
25	629	17.5	1010	US-08-882-046-7	Sequence 7, App1
26	629	17.5	1036	US-09-068-740A-6	Sequence 6, App1
27	629	17.5	1187	US-09-068-740A-7	Sequence 7, App1

28	629	17.5	1208	US-09-199-865-1	Sequence 1, App1
29	629	17.5	1218	US-08-400-159-6	Sequence 6, App1
30	629	17.5	1218	US-08-611-729A-6	Sequence 6, App1
31	629	17.5	1218	US-08-882-046-2	Sequence 2, App1
32	629	17.5	1218	US-09-068-740A-11	Sequence 11, App1
33	622	17.3	1238	US-09-214-278-5	Sequence 5, App1
34	619.5	17.2	1248	US-08-882-046-6	Sequence 6, App1
35	619	17.2	1248	US-09-214-278-7	Sequence 7, App1
36	616	17.1	1219	US-08-882-046-5	Sequence 5, App1
37	587	16.3	1404	US-08-400-159-2	Sequence 2, App1
38	587	16.3	1404	US-08-611-729A-2	Sequence 2, App1
39	560	15.6	3111	US-08-460-309-4	Sequence 4, App1
40	560	15.6	3111	US-08-125-077-4	Sequence 4, App1
41	556	15.4	833	US-08-264-534-6	Sequence 6, App1
42	556	15.4	833	US-08-083-590A-2	Sequence 2, App1
43	556	15.4	833	US-08-465-500-6	Sequence 6, App1
44	556	15.4	833	US-08-346-126-6	Sequence 6, App1
45	556	15.4	833	US-08-346-128-6	Sequence 6, App1

ALIGNMENTS

RESULT 1					
US-09-188-930-332					
; Sequence 332, Application US/09188930A					
; Patent No. 6150502					
; GENERAL INFORMATION:					
; APPLICANT: Watson, James D.					
; APPLICANT: Strachan, Iorna					
; APPLICANT: Sleeman, Matthew					
; APPLICANT: Orreust, Rene					
; APPLICANT: Murlson, James Greg					
; TITLE OF INVENTION: Compositions Isolated From Skin Cells					
; FILE REFERENCE: 11000.1011c1					
; CURRENT APPLICATION NUMBER: US/09/188,930A					
; NUMBER OF SEQ ID NOS: 348					
; SOFTWARE: FASTSEQ for Windows Version 3.0					
; SEQ ID NO 332					
; LENGTH: 299					
; TYPE: PRT					
; ORGANISM: Mouse					
US-09-188-930-332					
Query Match					
Best Local Similarity 39.9%; Pred. No. 4.4e-43;					
Matches 127; Conservative 27; Mismatches 118; Indels 46; Gaps 1;					
QY	165	GACCAAGFGRWGEDCDEGTGNDCHORCCONGATCCHVTEGCRCPGTGAFCEDL	224		
DB	4	GACCAAGFGRWGEDCDEGTGNDCHORCCONGATCCHVTEGCRCPGTGAFCEDL	63		
QY	225	CPPEKHPDCEORPCONGGVCHVTGECSCPSGMMGVGQPCPEGRFGKNCQEOCH	284		
DB	64	CPPEKHPDCEORPCONGGVCHVTGECSCPSGMMGVGQPCPEGRFGKNCQEOCH	123		
QY	285	NGGTDAATGSCCHSPYTERCDECPVGYGLCAETCCVNGRCYHVSAGCLCEAG	344		
DB	124	NGGTDAATGSCCHSPYTERCDECPVGYGLCAETCCVNGRCYHVSAGCLCEAG	183		
QY	345	PAGECARLCPESGLYIKDKRCPCHELENTSHCPMSGECACPKMSGLXNETSPGF	404		
DB	184	PAGECARLCPESGLYIKDKRCPCHELENTSHCPMSGECACPKMSGLXNETSPGF	197		
QY	405	YGEACQICSGONGADCVTGKTCAPFGKIDCSTPCLGTGNGCSRGKGNDAVC	464		
DB	198	YGEACQICSGONGADCVTGKTCAPFGKIDCSTPCLGTGNGCSRGKGNDAVC	257		
QY	465	SPVDGSCCTCRAGWGVDC	482		
DB	258	SPVDGSCCTCRAGWGVDC	275		

RESULT 2

US-09-188-930-192

Sequence 192, Application US/09188930A

Patent No. 6150502

GENERAL INFORMATION:

APPLICANT: Watson, James D.

APPLICANT: Strachan, Lorna

APPLICANT: Sleeman, Matthew

APPLICANT: Onrust, Rene

APPLICANT: Murlison, James Greg

TITLE OF INVENTION: Compositions Isolated From Skin Cells

FILE REFERENCE: 11000.1011c1

CURRENT APPLICATION NUMBER: US/09/188,930A

CURRENT FILING DATE: 1998-11-09

NUMBER OF SEQ ID NOS: 348

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 192

LENGTH: 299

TYPE: PRT

ORGANISM: mouse

FEATURE:

NAME/KEY: UNSURE

LOCATION: (98)..(98)

NAME/KEY: UNSURE

LOCATION: (239)...(239)

US-09-188-930-192

Query Match

Best Local Similarity 39.38; Pred. No. 8.4e-42;

Matches 125; Conservative 27; Mismatches 120; Indels 46; Gaps 1;

21.18; Score 759; DB 4; Length 299;

Best Local Similarity 39.38; Pred. No. 8.4e-42;

Matches 125; Conservative 27; Mismatches 120; Indels 46; Gaps 1;

21.18; Score 759; DB 4; Length 299;

Best Local Similarity 39.38; Pred. No. 8.4e-42;

Matches 125; Conservative 27; Mismatches 120; Indels 46; Gaps 1;

21.18; Score 759; DB 4; Length 299;

Best Local Similarity 39.38; Pred. No. 8.4e-42;

Matches 125; Conservative 27; Mismatches 120; Indels 46; Gaps 1;

21.18; Score 759; DB 4; Length 299;

Best Local Similarity 39.38; Pred. No. 8.4e-42;

Matches 125; Conservative 27; Mismatches 120; Indels 46; Gaps 1;

21.18; Score 759; DB 4; Length 299;

Best Local Similarity 39.38; Pred. No. 8.4e-42;

Matches 125; Conservative 27; Mismatches 120; Indels 46; Gaps 1;

21.18; Score 759; DB 4; Length 299;

Best Local Similarity 39.38; Pred. No. 8.4e-42;

Matches 125; Conservative 27; Mismatches 120; Indels 46; Gaps 1;

21.18; Score 759; DB 4; Length 299;

Best Local Similarity 39.38; Pred. No. 8.4e-42;

Matches 125; Conservative 27; Mismatches 120; Indels 46; Gaps 1;

21.18; Score 759; DB 4; Length 299;

Best Local Similarity 39.38; Pred. No. 8.4e-42;

Matches 125; Conservative 27; Mismatches 120; Indels 46; Gaps 1;

21.18; Score 759; DB 4; Length 299;

Best Local Similarity 39.38; Pred. No. 8.4e-42;

Matches 125; Conservative 27; Mismatches 120; Indels 46; Gaps 1;

21.18; Score 759; DB 4; Length 299;

Best Local Similarity 39.38; Pred. No. 8.4e-42;

Matches 125; Conservative 27; Mismatches 120; Indels 46; Gaps 1;

21.18; Score 759; DB 4; Length 299;

Best Local Similarity 39.38; Pred. No. 8.4e-42;

Matches 125; Conservative 27; Mismatches 120; Indels 46; Gaps 1;

21.18; Score 759; DB 4; Length 299;

Best Local Similarity 39.38; Pred. No. 8.4e-42;

Matches 125; Conservative 27; Mismatches 120; Indels 46; Gaps 1;

21.18; Score 759; DB 4; Length 299;

Best Local Similarity 39.38; Pred. No. 8.4e-42;

Matches 125; Conservative 27; Mismatches 120; Indels 46; Gaps 1;

21.18; Score 759; DB 4; Length 299;

Best Local Similarity 39.38; Pred. No. 8.4e-42;

Matches 125; Conservative 27; Mismatches 120; Indels 46; Gaps 1;

21.18; Score 759; DB 4; Length 299;

Best Local Similarity 39.38; Pred. No. 8.4e-42;

Matches 125; Conservative 27; Mismatches 120; Indels 46; Gaps 1;

21.18; Score 759; DB 4; Length 299;

Best Local Similarity 39.38; Pred. No. 8.4e-42;

Matches 125; Conservative 27; Mismatches 120; Indels 46; Gaps 1;

21.18; Score 759; DB 4; Length 299;

Best Local Similarity 39.38; Pred. No. 8.4e-42;

ADDRESSEE: PENNIE & EDMONDS

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION NUMBER: US/08/185,432

APPLICATION NUMBER: US/08/185,432

FILING DATE: 21-JAN-1994

CLASSIFICATION: 530

ATTORNEY/AGENT INFORMATION:

NAME: Mistrock, S. Leslie

REGISTRATION NUMBER: 18,872

REFERENCE/DOCKET NUMBER: 7326-006

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 790-9090

TELEFAX: (212) 869-8864/9741

TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 18:

SEQUENCE CHARACTERISTICS:

LENGTH: 2523 amino acids

TYPE: amino acid

TOPOLOGY: unknown

MOLECULE TYPE: protein

US-08-185-432-18

Query Match

Best Local Similarity 20.0%; Score 719; DB 1; Length 2523;

Matches 222; Conservative 60; Mismatches 222; Indels 355; Gaps 50;

20.0%; Score 719; DB 1; Length 2523;

Matches 222; Conservative 60; Mismatches 222; Indels 355; Gaps 50;

20.0%; Score 719; DB 1; Length 2523;

Matches 222; Conservative 60; Mismatches 222; Indels 355; Gaps 50;

20.0%; Score 719; DB 1; Length 2523;

Matches 222; Conservative 60; Mismatches 222; Indels 355; Gaps 50;

20.0%; Score 719; DB 1; Length 2523;

Matches 222; Conservative 60; Mismatches 222; Indels 355; Gaps 50;

20.0%; Score 719; DB 1; Length 2523;

Matches 222; Conservative 60; Mismatches 222; Indels 355; Gaps 50;

20.0%; Score 719; DB 1; Length 2523;

Matches 222; Conservative 60; Mismatches 222; Indels 355; Gaps 50;

20.0%; Score 719; DB 1; Length 2523;

Matches 222; Conservative 60; Mismatches 222; Indels 355; Gaps 50;

20.0%; Score 719; DB 1; Length 2523;

Matches 222; Conservative 60; Mismatches 222; Indels 355; Gaps 50;

20.0%; Score 719; DB 1; Length 2523;

Matches 222; Conservative 60; Mismatches 222; Indels 355; Gaps 50;

20.0%; Score 719; DB 1; Length 2523;

Matches 222; Conservative 60; Mismatches 222; Indels 355; Gaps 50;

20.0%; Score 719; DB 1; Length 2523;

Matches 222; Conservative 60; Mismatches 222; Indels 355; Gaps 50;

20.0%; Score 719; DB 1; Length 2523;

Matches 222; Conservative 60; Mismatches 222; Indels 355; Gaps 50;

20.0%; Score 719; DB 1; Length 2523;

Matches 222; Conservative 60; Mismatches 222; Indels 355; Gaps 50;

20.0%; Score 719; DB 1; Length 2523;

Matches 222; Conservative 60; Mismatches 222; Indels 355; Gaps 50;

20.0%; Score 719; DB 1; Length 2523;

Matches 222; Conservative 60; Mismatches 222; Indels 355; Gaps 50;

20.0%; Score 719; DB 1; Length 2523;

Matches 222; Conservative 60; Mismatches 222; Indels 355; Gaps 50;

20.0%; Score 719; DB 1; Length 2523;

Matches 222; Conservative 60; Mismatches 222; Indels 355; Gaps 50;

20.0%; Score 719; DB 1; Length 2523;

Matches 222; Conservative 60; Mismatches 222; Indels 355; Gaps 50;

20.0%; Score 719; DB 1; Length 2523;

Matches 222; Conservative 60; Mismatches 222; Indels 355; Gaps 50;

20.0%; Score 719; DB 1; Length 2523;

Matches 222; Conservative 60; Mismatches 222; Indels 355; Gaps 50;

20.0%; Score 719; DB 1; Length 2523;

Matches 222; Conservative 60; Mismatches 222; Indels 355; Gaps 50;

20.0%; Score 719; DB 1; Length 2523;

Matches 222; Conservative 60; Mismatches 222; Indels 355; Gaps 50;

20.0%; Score 719; DB 1; Length 2523;

Matches 222; Conservative 60; Mismatches 222; Indels 355; Gaps 50;

20.0%; Score 719; DB 1; Length 2523;

Matches 222; Conservative 60; Mismatches 222; Indels 355; Gaps 50;

20.0%; Score 719; DB 1; Length 2523;

```

Db 1069 NGKCMQJTNFNYKCECKSGMTGYCDVPSVCEVAQAQGVDIYHLCRNSGMVDGTGNTH 1128
QY 443 ---CPLGTGYNCSRR---CG---CKNDAYSPVDG---SCTCKAGMHGVDGS----- 483
Db 1129 FCRQAGTGTGCEQVDECPNPNQONATCTDYLGGISCECVAGHYVKNSEELNCLTS 1188
QY 484 -----IRCPSTGWFGCNLT---C-----OCLNGAGCNTLDG 512
Db 1189 HPCQNGGTCTDLIMTYKSCPRGTGVHCEINVDCTPFYDSFTLEPRCFNNKCIDRVG 1248
QY 513 ---TCTCARGMGEKCE-----LPCQD---GTGLNCAE-----RCDC-----SH 546
Db 1249 GYNCTCPRGFEVGERCEGDVNECLSNPCDSRGTQ---NCTQVNDYRCCECRGFTGRCESV 1306
QY 547 ADGC-----HPTTGH---CRCLPGMSGVHCD----- 569
Db 1307 VDGCKMPCNRGTCNAVANSNERGFICKCPGFGDGTCEYDSRCSNLRQNGGTCTISVL 1366
QY 570 ---SVCAGRWGPNCSLP 584
Db 1367 TSSKVCSEGYTGATCQYP 1385

```

```

RESULT 4
US-08-899-232-3
; Sequence 3, Application US/08899232
; Patent No. 6436650
; GENERAL INFORMATION:
; APPLICANT: Artavanis-Tsakonas, Spyridon
; APPLICANT: Qi, Huijin
; TITLE OF INVENTION: ACTIVATED FORMS OF NOTCH AND METHODS BASED THEREON
; FILE REFERENCE: 7326-046
; CURRENT APPLICATION NUMBER: US/08/899,232
; CURRENT FILING DATE: 1997-07-23
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 3
; LENGTH: 2523
; TYPE: PRT
; ORGANISM: Xenopus sp.
US-08-899-232-3

```

```

Query Match 20.0%; Score 719; DB 4; Length 2523;
Best Local Similarity 25.8%; Pred. No. 2e-38;
Matches 222; Conservative 60; Mismatches 222; Indels 355; Gaps 50;

```

```

QY 5 LNSLSLFLCL-----LICHWIGTASPLLED-----PNVCSHW-----EYS 41
Db 603 INELSKRCLNGGQCTDRENGYICTCPKGTGVNCEFKRIDCASNLCDNGKCIDKIDYE 662
QY 42 VYQESYPHFPDQIYTT-----SCTDIIMWFCTHRVSYRTAYRHGKTYRR 90
Db 663 CTCEPGTGTGKCLNINIBCDSPNCRNGSTCKADQJNGFTCV----- 702
QY 91 KSGCCPGFYESGEMC---VPHC-ADKCVHGR---IAPNTQCEPBGWGTNC---SSACD 140
Db 703 ---CPDGYHD-HMCLSEVNECNSNPCHGACHDGVANGKDCDEAGMSGSCDINNCE 757
QY 141 GDHNGPHCTSRCKCNALCPITGA--CHCAAFRGHRCRDREOGYTGDCHQRCQ 198
Db 758 SN-----PCMNGGTCKDMGTGAYICTKAGFSGPNCO-----TINIECSSN-PCL 800
QY 199 NGATC-DHVTG-ECRCPPGYTGAFCEDLCPRGKGPCEQRC---PCONGVCHH---V 249
Db 801 NHGTCIDVAGYKCNCHLPITGALCEAVLAP-----CASSPCKNGRCKESEDFE 850
QY 250 TGECSGPGWGTVCQPCPRGFRGKNCSDC---QCHNGTCDATG--QCHSPGYTG 304
Db 851 TFSCECPGMOGOTC-----EIDMNECVNRPCKNGATCQNTNGSYKCNCKPGYTG 900
QY 305 ERQ---DECPVGTGYVLCATCCQVANGKGYHSGA--CLCEAGFAGEREALD----- 354
Db 901 RNCMDIDDC-----QPNPCHHGSGSDSINMFECNCPAGFRGPKCEDIEDINCCAS 950

```

```

QY 355 -----CEPHYGIKDRRCPECHLE-----NTHSCHPMGSG-----ECAC 387
Db 951 NPKNGANGTCQVNSYNTCTCPGFSGICEBSNTDCTSSCFNNGTC--IDGINTFTQCQ 1008
QY 388 KPGWSGLYC---NE-----TCSPGFYBACQOI---CS---CQ 416
Db 1009 PPGFTGASYCOHDINECKSPCLANGTCQDSYGYKCTCPQGYTGAINCNLVRMCDSSPCK 1068
QY 417 NGACCDSVTG--KCTCARGFEGIDCSFP----- 442
Db 1069 NGKCMQJTNFNYKCECKSGMTGYCDVPSVCEVAQAQGVDIYHLCRNSGMVDGTGNTH 1128
QY 443 ---CPLGTGYNCSRR---CG---CKNDAYSPVDG---SCTCKAGMHGVDGS----- 483
Db 1129 FCRQAGTGTGCEQVDECPNPNQONATCTDYLGGISCECVAGHYVKNSEELNCLTS 1188
QY 484 -----IRCPSTGWFGCNLT---C-----OCLNGAGCNTLDG 512
Db 1189 HPCQNGGTCTDLIMTYKSCPRGTGVHCEINVDCTPFYDSFTLEPRCFNNKCIDRVG 1248
QY 513 ---TCTCARGMGEKCE-----LPCQD---GTGLNCAE-----RCDC-----SH 546
Db 1249 GYNCTCPRGFEVGERCEGDVNECLSNPCDSRGTQ---NCTQVNDYRCCECRGFTGRCESV 1306
QY 547 ADGC-----HPTTGH---CRCLPGMSGVHCD----- 569
Db 1307 VDGCKMPCNRGTCNAVANSNERGFICKCPGFGDGTCEYDSRCSNLRQNGGTCTISVL 1366
QY 570 ---SVCAGRWGPNCSLP 584
Db 1367 TSSKVCSEGYTGATCQYP 1385

```

```

RESULT 5
US-08-400-159-10
; Sequence 10, Application US/08400159
; Patent No. 5869282
; GENERAL INFORMATION:
; APPLICANT: Ish-Horowitz, David
; APPLICANT: Henriques, Domingos M.P.
; APPLICANT: Lewis, Julian H.
; APPLICANT: Myatt, Anna M.
; APPLICANT: Fleming, Robert J.
; APPLICANT: Artavanis-Tsakonas, Spyridon
; APPLICANT: Mann, Robert S.
; TITLE OF INVENTION: NUCLEOTIDE AND PROTEIN SEQUENCES OF THE
; TITLE OF INVENTION: SERRATE GENE AND METHODS BASED THEREON
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/400,159
; FILING DATE: 07-MAR-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7326-029
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864

```


QY 428 CTCAPGFKGIDC-----STPCPLGTYGINCSS--RCGK-----NDAYCS 465
 DB 668 CVCSPGFTGQRCNIDIDECASNPCKRGATCINGVNGFRICPEBPHHPSYQVNECLSN 727
 QY 466 P-VGSGCT-----CKAGMHGVDCSI-----RCPSGT 490
 DB 728 PCINGNCTGSLGYKCLDADGAVGINCEDVKNELSNPCQNGTCDNLVNGYRGTCKKGF 787
 QY 491 WGFQCNLTQ-----CLNGGAC-----507
 DB 788 KGYNQVVIDCASNPLCNOGTCEDDISGYTCHCVLPYTGKNCQTVLAPCSPNCEANAY 847
 QY 508 -----NTLDGTCAPGMRGKCEL-----PCOD-----GTV-----GLN 537
 DB 848 CKESPNEFSTYCLCAPGMOGRCITIDIDECISKPCMNHGLCHNTGSGYMCECPGFSGMD 907
 QY 538 CAEKCD-----CSHADGCHP--TTGHCRCLPGMSGVCHDSYCAEGRMKPRC-SLPC 585
 DB 908 CEEDIDCLANPCQNGSGCMDGVNTFSCCLPFGTGDKCQTDNME-----CLSEPC 958

RESULT 10

US-08-083-590A-19
 ; Sequence 19, Application US/08083590A
 ; Patent No. 5786158
 ; GENERAL INFORMATION:
 ; APPLICANT: Artavanis-Tsakonas, S. et al.
 ; TITLE OF INVENTION: Therapeutic And Diagnostic Methods
 ; TITLE OF INVENTION: And Compositions Based On No. 5786158ch Proteins And
 ; TITLE OF INVENTION: Nucleic Acids
 ; NUMBER OF SEQUENCES: 21.
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Pennie & Edmonds
 ; STREET: 1155 Avenue of the Americas
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: U.S.A.
 ; ZIP: 10036
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/083,590A
 ; FILING DATE: 25-JUN-1993
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Mistrock, S. Leslie
 ; REGISTRATION NUMBER: 18,872
 ; REFERENCE/DOCKET NUMBER: 7326-015
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 212 790-9090
 ; TELEFAX: 212 8698864/9741
 ; TELEX: 66141 PENNIE
 ; INFORMATION FOR SEQ ID NO: 19:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2471 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: unknown
 ; MOLECULE TYPE: peptide
 ; US-08-083-590A-19

Query Match 18.5%; Score 666.5; DB 1; Length 2471;

Best Local Similarity 23.3%; Pred. NO. 4.5e-35;

Matches 209; Conservative 64; Mismatches 191; Indels 433; Gaps 46;

QY 93 QCCPGFYEGEMVPHCADKCVHGR-CIAPNTCO-----CEPGMGSTNCSACDG 141
 DB 91 RCASGF--TGEDQCYTSHPCFVSRPCLNGSTCHMLSDTYECTCQVGFTEKEC-----142
 QY 142 DHMGPHCTSRQCKNGALCNPTIG--ACHCAAGFRGWRCEDRCEQGTGYGNDCHORCOQN 199

DB 143 -QWTDACLSH-PCANGSTCTTVANQFSCKLGTGQCE-----TDVNCDDIPHQH 194
 QY 200 GATCDHTYG--ECRCPPGYTAFCEDL-----CPGKHG 231
 DB 195 GGCTILNPGSYQCCPOGFTGQYCDSLYVPCASPVCYNGTCRQYGTFTPCNCLPFGEG 254
 QY 232 PQCEQ-----RCP-----CONGV 245
 DB 255 STCERNIDDCPNHRCQNGVCVVDGVNTYNCRCPPQMTGQFCTEDYDECLLPNMCQNGT 314
 QY 246 CHAVTG--ECSPPSGMWGTVCQ-----PCEBGRGKNC- 277
 DB 315 CANRNGGYGVYVNGMSDDCSENIIDCAFASCTPGSTCIDRVASFSCMCPEGRKALLCH 374
 QY 278 -SQEC--QCHNGTCA--ATGQ--CHCSGYTGERQ--DECPVGTGYVLAETQC 326
 DB 375 LDDACISNPKRGALCDTNPLNGYITCPCPGYRGADCTEDYDECBAM-----ANSNP 427
 QY 327 VNGKCYHVSQA--CLCEAGFAGERCE-----ARLCEGLY 360
 DB 428 EHAKCYVTDGAFHCECLKGTAGRCENDINECHSDPCQNDATCLDKIGFTCLCMQFK 487
 QY 361 GIKCDKR-----CP-----CHLE-----NTHSC 378
 DB 488 GVHCELEINECOSNPNVCVNGQCVDKVNRFOCLCPGFGPYCQIDIDDCSSTPLNGAKC 547
 QY 379 --HMSGCAKPRMSGLYCNET-----CSRFYEAC-QQ 411
 DB 548 IDHNGYECQCATGTFGLCEBENIDNCDPDCHHGQCODGIDSYTYCICNPGYMAICSDQ 607
 QY 412 I-----CSCQNGA-----DCDS--VTG-----K 427
 DB 608 IDECYSSPLNDGRCLDLYNGYQCNPGTSGVNCNEINFGDCASNPLCHIGICMGINRYS 667
 QY 428 CTCAPGFKGIDC-----STPCPLGTYGINCSS--RCGK-----NDAYCS 465
 DB 668 CVCSPGFTGQRCNIDIDECASNPCKRGATCINGVNGFRICPEBPHHPSYQVNECLSN 727
 QY 466 P-VGSGCT-----CKAGMHGVDCSI-----RCPSGT 490
 DB 728 PCINGNCTGSLGYKCLDADGAVGINCEDVKNELSNPCQNGTCDNLVNGYRGTCKKGF 787
 QY 491 WGFQCNLTQ-----CLNGGAC-----507
 DB 788 KGYNQVVIDCASNPLCNOGTCEDDISGYTCHCVLPYTGKNCQTVLAPCSPNCEANAY 847
 QY 508 -----NTLDGTCAPGMRGKCEL-----PCOD-----GTV-----GLN 537
 DB 848 CKESPNEFSTYCLCAPGMOGRCITIDIDECISKPCMNHGLCHNTGSGYMCECPGFSGMD 907
 QY 538 CAEKCD-----CSHADGCHP--TTGHCRCLPGMSGVCHDSYCAEGRMKPRC-SLPC 585
 DB 908 CEEDIDCLANPCQNGSGCMDGVNTFSCCLPFGTGDKCQTDNME-----CLSEPC 958

RESULT 11

US-08-532-384-19
 ; Sequence 19, Application US/08532384
 ; Patent No. 6083904
 ; GENERAL INFORMATION:
 ; APPLICANT: Artavanis-Tsakonas, S. et al.
 ; TITLE OF INVENTION: Therapeutic And Diagnostic Methods
 ; TITLE OF INVENTION: And Compositions Based On No. 6083904ch Proteins And
 ; TITLE OF INVENTION: Nucleic Acids
 ; NUMBER OF SEQUENCES: 21
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Pennie & Edmonds
 ; STREET: 1155 Avenue of the Americas
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: U.S.A.
 ; ZIP: 10036

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/532,384
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/083,590
FILING DATE: 25-JUN-1993
ATTORNEY/AGENT INFORMATION:
NAME: MISTOCK, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7326-015
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 790-9090
TELEFAX: 212 8698864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 2471 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-532-384-19

Query Match      18.5%; Score 666.5; DB 3; Length 2471;
Best Local Similarity 23.3%; Pred. 4.5e-35;
Matches 209; Conservative 64; Mismatches 191; Indels 433; Gaps 46;

OY 93 OCCPGFYESGEMCVPHCADKCYHGR-CIAPNTCQ-----CEPGMGTCSSACDG 141
DB 91 RCASGF--TGEDCOYSTHPCFVSRLCLNGTCHMLSRDYTECTOVFTGKEC----- 142
OY 142 DHMGPHCTSRCCCKNGALCNPTIG--ACHCAGFRGMRCEDEGCTYNDCHORCCON 199
DB 143 -QMTDACLSH-PCANGSTTTVANOFSCKLTGFTGKCE-----TDVNECDIRGHCOH 194
OY 200 GATCDHYTG--ECRCPPGTGAFCEDL-----CPRGKHG 231
DB 195 GGTCLNLPESYQCQCQPGTGOYSLYVPCASPVCVNGGTGRGTGDFTECNCLPFGFG 254
OY 232 POCCEQ-----RCP-----CONGV 245
DB 255 STCERNIDDCPNHRCONGVCVDGVNTYNCRCRPQWGTGQCTEDYDECLLOPNACONGST 314
Y 246 CHHYTG--ECSCPSGMGMGVCGQ-----PCPRGRGKNC- 277
DB 315 CANRRNGYGCVCVNGNSGDSCSENIDDCAFASCTPGSTCIDRVASFSCMCPGRKAGLLCH 374
OY 278 -SQEC---QCHNGGTCD--ATGQ--CHCSPGYTGERCO---DECPVGTGYVLAETCQC 326
DB 375 LDDACISNCHKGALCDTRPLNGOYITCPOGYTKADCTEDYDECAM-----ANSNPC 427
OY 327 VNGKCYHYVSGA--CLCEAGFAGERCE-----ARLCEPGLY 360
DB 428 EHAGKCVNTDGAHFHCECLLGAYAGPRCEMDINECHSDPCONDATCLDKIGFTCLAMPGRK 487
OY 361 GIKCDKR-----CP-----CHLE-----NTHSC 378
DB 488 GVHCELEINECOSNPCVNNOCVDKYNRFQCLCPGFTGTPVCOIDIDDCSSPTCLNGARC 547
OY 379 --HPMSGECACRPMGSLYCNET-----CSPGYGEMAC--QO 411
DB 548 IDHPNGYEQCATFTGTGVLCEENIDNCBPPCHHGOCODGIDSYICICPMGTGALCSQ 607
OY 412 I-----CSCQNGA-----DCDS---VTG-----K 427
DB 608 IDECYSSPLNDGRCDLNVGYQCNCQPGTSGVNCINPDCAASNPCINHGICMDGINRS 667

```

```

OY 428 CTCARQFKRIDC-----STPCPLGTGYINCSS--RCGCK-----NDAVCS 465
DB 668 CVCSPGFTGQRCNIDIDECASNPCKRGATCINGVNGFRICBEPHHPSCYSQVNECLSN 727
OY 466 P-VDSCT-----CKAGHGVDCSI-----RCPST 490
DB 728 PCIRHNCITGGLSGYKCLDAGVNGINCEYDKNDECLSNPCQNGTCDNLVNGYRCKKGF 787
OY 491 WFGGCLNTCQ-----CLNGGAC----- 507
DB 788 KGVNCOVINIDECASNPCLNMGTCFDDISGTYCHVLPYTKNCKQYVLAPCSNPENNAV 847
OY 508 -----NTLDGTCTCARPMRGEKCEL-----PCOD-----GTY-----GIN 537
DB 848 CKESPFESYTCICAPMGWGORCTIDIDECISKPCMNHGLCHNTGGSYMCCECPPEFGSD 907
OY 538 CAERCD-----CSHADGCHP--TTGHCRCLPGMSGVHCHDSYCARGRMPNC-SLPC 585
DB 908 CEDDIDDCIANPCQNGGSCMDGVNTFSCICLPLGFTGDKCQTDME-----CLSEPC 958

RESULT 12
US-08-899-232-1
; Sequence 1, Application US/08899232
; Patent No. 6436650
; GENERAL INFORMATION:
; APPLICANT: Artavanis-Tsakonas, Spyridon
; TITLE OF INVENTION: ACTIVATED FORMS OF NOTCH AND METHODS BASED THEREON
; FILE REFERENCE: 7326-046
; CURRENT APPLICATION NUMBER: US/08/899,232
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2471
; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-899-232-1

Query Match      18.5%; Score 666.5; DB 4; Length 2471;
Best Local Similarity 23.3%; Pred. 4.5e-35;
Matches 209; Conservative 64; Mismatches 191; Indels 433; Gaps 46;

OY 93 OCCPGFYESGEMCVPHCADKCYHGR-CIAPNTCQ-----CEPGMGTCSSACDG 141
DB 91 RCASGF--TGEDCOYSTHPCFVSRLCLNGTCHMLSRDYTECTOVFTGKEC----- 142
OY 142 DHMGPHCTSRCCCKNGALCNPTIG--ACHCAGFRGMRCEDEGCTYNDCHORCCON 199
DB 143 -QMTDACLSH-PCANGSTTTVANOFSCKLTGFTGKCE-----TDVNECDIRGHCOH 194
OY 200 GATCDHYTG--ECRCPPGTGAFCEDL-----CPRGKHG 231
DB 195 GGTCLNLPESYQCQCQPGTGOYSLYVPCASPVCVNGGTGRGTGDFTECNCLPFGFG 254
OY 232 POCCEQ-----RCP-----CONGV 245
DB 255 STCERNIDDCPNHRCONGVCVDGVNTYNCRCRPQWGTGQCTEDYDECLLOPNACONGST 314
Y 246 CHHYTG--ECSCPSGMGMGVCGQ-----PCPRGRGKNC- 277
DB 315 CANRRNGYGCVCVNGNSGDSCSENIDDCAFASCTPGSTCIDRVASFSCMCPGRKAGLLCH 374
OY 278 -SQEC---QCHNGGTCD--ATGQ--CHCSPGYTGERCO---DECPVGTGYVLAETCQC 326
DB 375 LDDACISNCHKGALCDTRPLNGOYITCPOGYTKADCTEDYDECAM-----ANSNPC 427
OY 327 VNGKCYHYVSGA--CLCEAGFAGERCE-----ARLCEPGLY 360
DB 428 EHAGKCVNTDGAHFHCECLLGAYAGPRCEMDINECHSDPCONDATCLDKIGFTCLAMPGRK 487
OY 361 GIKCDKR-----CP-----CHLE-----NTHSC 378

```


SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 2
 LENGTH: 2556
 TYPE: PR
 ORGANISM: Homo sapiens
 US-08-899-232-2

Query Match 18.4%; Score 663; DB 4; Length 2556;
 Best Local Similarity 26.7%; Pred. No. 7.8e-35;
 Matches 204; Conservative 44; Mismatches 192; Indels 324; Gaps 49;

94 CCGFYESEMCVPHCADCVHGR-----IAPTCOCPEPMGGTSSACDGDH 143
 DB 89 CALGF--SGPLCLPLDNACLTNPCRNGGTCDLLTLEKRCRCRPPMSGSKCOA----- 141
 144 MPRHCTSRCCCKGALCNITGA--CHCAAGFRG--WPCEDRCCEG--TYGNCQHO- 193
 DB 142 --DPCASN-PCANGGCLPEFASITICHCPSPFPGTCMDVNECGKPRLCRHGTCHNE 198
 194 ----RC-----OCNGATC--DHVTGECRCPPGYTAFCE-- 222
 DB 199 VGSYRCVCAATHGPRCEPVPYPCSPPCONGGTCHPTGDVTHEACALGFTGQNCENI 258
 223 DLCPFG--KHGPOC-----EQRCP-----CONGVCHHTG- 251
 DB 259 DDCPGNNCKNGACVGVNTYVNCPCPEWTGYCTEDVDECOLMPNACONGTCHHTHG 318
 252 --ECSPSGMNGVYCGO-----PCPEGRFGKNC--SQEC-- 281
 DB 319 YNCVAVNGMTGECSENIDDCASACFHATCHDRVASFYCECPHRTGLCHLNACIS 378
 282 --OCHNGTCD--ATGO--CHCSPGYTGERCQ--DECPVGYGVLAETCQCQVNGKCY 333
 DB 379 NPCNENSGNDITPNVNGKALCTCPSPGYTGPACSDVDECSIGAN-----PCENHAKCI 430
 334 HVSGA--CLCEAGFAGEGCEARLCPEGLGKIDKRC--PCHLENTHSCHPMSGE--CA 386
 DB 431 NTLGSECCCLGTYGTPRCEIDV-----NECVSNPC--QNDATCLDQIGEFQCM 477
 387 CRPGWSGLYC-----NE--TCSPGFYGERCQ--QICS--C 415
 DB 478 CMPEGEVHCEVNTDECASSPCLHNGRCLDKINEFQCEPPTGTHLCOYDVECASTPC 537
 416 QNGADC--DSVTKCTCAGFGKIDCT-----PCPLGTYGKNCSSRCGCKNDAYCS 465
 DB 538 KNGAKLDGPNYTYCTCTGTYGTCEVDIDCEDDPCHYGS-----CK-DGVAT 586
 466 PVDSCTCKAGMHGVNCSI--RC--PSGTMGFGCNLCQ-----CLNG-- 504
 DB 587 --FTCLCARPTTGHHCETNINECSSQPCRLMG-----TCDDPDNAYLCFCLGTTGPMC 638
 505 -----GAC-NTLDG--TCTCAPGWRGKCELT-----PCODGTY 534
 DB 639 EINALDCASSPDCSGTCLDKIDYECACEGYTSGMNSNIDECAGNPGCHNGTCDGJN 698
 535 GLNCAERC-----DCSHADGCHPTTGH--CRCLPGMSGVHCD----- 569
 DB 699 GFTC--RCPEGYHPTCLSEVNECNSNPGVHGACMDSLNGKCDPCBPSGTCNDINNNE 756
 570 -----SYCAEGRMGPN-----SLPC 585
 DB 757 CESNPCVNGTCKDMTSGIVCTCMBESGPNCTNINECASNPC 800

RESULT 15
 US-08-083-590A-20
 Sequence 20, Application US/08083590A
 Patent No. 5786158
 GENERAL INFORMATION:
 APPLICANT: Artavanis-Tsakonas, S. et al.
 TITLE OF INVENTION: Therapeutic And Diagnostic Methods
 TITLE OF INVENTION: And Compositions Based On No. 5786158ch Proteins And
 TITLE OF INVENTION: Nucleic Acids

NUMBER OF SEQUENCES: 21
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Pennie & Edmonds
 STREET: 1155 Avenue of the Americas
 CITY: New York
 STATE: New York
 COUNTRY: U.S.A.
 ZIP: 10036
 COMPUTER READABLE FORM:
 MEDIUM TYPE: floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/083,590A
 FILING DATE: 25-JUN-1993
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Mistrock, S. Leslie
 REGISTRATION NUMBER: 18,872
 REFERENCE/DOCKET NUMBER: 7326-015
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212 790-9090
 TELEFAX: 212 869864/9741
 TELEX: 66141 PENNIE
 INFORMATION FOR SEQ ID NO: 20:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2556 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: unknown
 MOLECULE TYPE: peptide
 US-08-083-590A-20

Query Match 18.4%; Score 662; DB 1; Length 2556;
 Best Local Similarity 26.3%; Pred. No. 9e-35;
 Matches 201; Conservative 46; Mismatches 194; Indels 322; Gaps 46;

94 CCGFYESEMCVPHCADCVHGR-----IAPTCOCPEPMGGTSSACDGDH 143
 DB 89 CALGF--SGPLCLPLDNACLTNPCRNGGTCDLLTLEKRCRCRPPMSGSKCOA----- 141
 144 MPRHCTSRCCCKGALCNITGA--CHCAAGFRGCE--DRCEG--TYGNCQHO- 193
 DB 142 --DPCASN-PCANGGCLPEFASITICHCPSPFPGTCMDVNECGKPRLCRHGTCHNE 198
 194 ----RC-----OCNGATC--DHVTGECRCPPGYTAFCE-- 222
 DB 199 VGSYRCVCAATHGPRCEPVPYPCSPPCONGGTCHPTGDVTHEACALGFTGQNCENI 258
 223 DLCPFG--KHGPOC-----EQRCP-----CONGVCHHTG- 251
 DB 259 DDCPGNNCKNGACVGVNTYVNCPCPEWTGYCTEDVDECOLMPNACONGTCHHTHG 318
 252 --ECSPSGMNGVYCGO-----PCPEGRFGKNC--SQEC-- 281
 DB 319 YNCVAVNGMTGECSENIDDCASACFHATCHDRVASFYCECPHRTGLCHLNACIS 378
 282 --OCHNGTCD--ATGO--CHCSPGYTGERCQ--DECPVGYGVLAETCQCQVNGKCY 333
 DB 379 NPCNENSGNDITPNVNGKALCTCPSPGYTGPACSDVDECSIGAN-----PCENHAKCI 430
 334 HVSGA--CLCEAGFAGEGCEARLCPEGLGKIDKRC--PCHLENTHSCHPMSGE--CA 386
 DB 431 NTLGSECCCLGTYGTPRCEIDV-----NECVSNPC--QNDATCLDQIGEFQCM 477
 387 CRPGWSGLYC-----NE--TCSPGFYGERCQ--QICS--C 415
 DB 478 CMPEGEVHCEVNTDECASSPCLHNGRCLDKINEFQCEPPTGTHLCOYDVECASTPC 537
 416 QNGADC--DSVTKCTCAGFGKIDCT-----PCPLGTYGKNCSSRCGCKNDAYCS 465
 DB 538 KNGAKLDGPNYTYCTCTGTYGTCEVDIDCEDDPCHYGS-----CK-DGVAT 586

```
OY 466 PYDGSCTCKAGWHGVDCSIR-----CPSGTWFGCNL--- 497
      :|:|:|
DB 587 ---FTCLCRPGYTGHHCEFININECSSQPCRLNGTCODPDNAYLCFLKGTGPNCIIND 643
      :|:|:|
OY 498 ---TCCLNGGACNTLDG-TCTCAPGWRGECLE-----PCODGTGLNCA 539
      :|:|:|
DB 644 DCASSPCDSTCLDKIDGYECACEPGYTGSMCNSNIDECAGNPCHNGGTCEDEGNGFTC- 702
      :|:|:|
OY 540 ERDCDSHADGCHPTT-----GHCR-----CLPGMSGVHCD----- 569
      :|:|:|
DB 703 -RC---PEGYHDPYCLSEVNECNSNPCVHGACRDSLNGYKCCDDPGWMSGTNCIDINNNEC 757
      :|:|:|
OY 570 -----SVCABGRWGPNC-----SLPC 585
      :|:|:|
DB 758 ESNPCVNGGTCKDMTSGIYCTCREGFSGPNCOTNINECASNPC 800
      :|:|:|
```

Search completed: May 9, 2003, 11:59:33
Job time : 24.5435 secs

THIS PAGE BLANK (USPTO)

OY 351 EARLCEGLYGIKDKRCPCHLENTHS 377
DB 249 EARLCEGLYGIKDKRCPCHLENTHS 275

RESULT 3

ABG08033
ID ABG08033 standard; Protein; 878 AA.

AC ABG08033;

DT 13-FEB-2002 (first entry)

DE Novel human diagnostic protein #8024.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

BN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI: 2001-639362/73.
DR N-PSDB: AAS72220.

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -

PS Claim 20; SEQ ID No 38392; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcc_sequences.

XX Sequence 878 AA;

Query Match 36.4%; Score 1309; DB 22: Length 878;
Best Local Similarity 86.9%; Pred No. 5.5e-55;
Matches 218; Conservative 3; Mismatches 14; Indels 16; Gaps 3;

OY 338 ACLCEAGFAGERCARLCEGLYGIKDKRCPCHLEN--THSCHPMSEGCACKPGWSGLY 395

DB 5 ALLCQUTVA-----C-----ISADLICFPANEOQLVACHHPMSGCAKCPGWSGLY 50
OY 396 CNETSGPGFYGGACQOICSCQNGADCDSTYKCTCAPGKIDCSTPCPLGTGYNCSR 455
DB 51 CNETSGPGFYGGACQOICSCQNGADCDSTYKCTCAPGKIDCSTPCPLGTGYNCSR 110
OY 456 CGCKNDAYCSPVDGSCCTKAGWHGVDCSTRCPSGTWFGFCNLTCOC LINGACNTLDGTCT 515
DB 111 CGCKNDAYCSPVDGSCCTKAGWHGVDCSTRCPSGTWFGFCNLTCOC LINGACNTLDGTCT 170
OY 516 CAPGMRGECCELPQDGTGTGLNCAERCDCSHADGCHPTTGHCRCPLPGMSGVHCDVCAEG 575
DB 171 CAPGMRGECCELPQDGTGTGLNCAERCDCSHADGCHPTTGHCRCPLPGMSGVHCDVCAEG 230
OY 576 RWGPNCSLPCY 586
DB 231 RWGPNCSLPCY 241

RESULT 4

ABG22559
ID ABG22559 standard; Protein; 466 AA.

AC ABG22559;

DT 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #22550.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

BN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI: 2001-639362/73.
DR N-PSDB: AAS86746.

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -

PS Claim 20; SEQ ID No 52918; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity

CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABC00010-ABC30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at [ftp://wipo.int/pub/published_pcr_sequences](http://wipo.int/pub/published_pcr_sequences).

5Q Sequence 466 AA;

Query Match	35.3%	Score 1272;	DB 22,	Length 466;
Best Local Similarity	76.9%	Pred. No. 2e-53;		
Matches 220; Conservative	3;	Mismatches 23;	Indels 40;	Gaps 5.

QY	118	CIANTCCCEEGMGATGNCSSACDDBDHMGPHCTSRCCCKNGLCPRTGACHC--AAGRGM	176
Db	39	CFSP---QSGPAAVITGVPTPAAG-----ASAKMGLCAPIPTIGACPLMLRGFRGM	85
QY	177	RCEDRCEDGTYGNDCHQRCCQNGATCDHYTGECRCPPGYTGAFCEDLCPGKHKPCQ	236
Db	86	RCEBRCBEGTGNCHORCCQNGATCDHYTGECRCPPGYTGAFCEDLCPGKHKPCQ	145
QY	237	RCPCONGVCHHVTGECSCPSGNN-----GVNCSQCPCEBREFGNCS	278
Db	146	RCPCONGVCHHVTGECSCPSGNNLSPFGMRPIFXSLMGGVTCQCPCEBEGFGNCS	205
QY	279	QECCHNGGTCDAAATGCHCSPGYTGR-----QDEBCPVGTGYVLCAETTCOCYNG	330
Db	206	QECCHNGGTCDAAATGCHCSPGYTGERAAVPDVRCQDEBCPVGTGYVLCAETTCOCYNG	265
QY	331	KCYHVSACLCCEAGFAGERCBARLCPEGLGKCDKRCPHLENTH	376
Db	266	KCYHVSACLCCEAGFAGERCBARLCPEGLGKCDKRCPHLENTH	311

RESULT 5
ABB66756
ID ABB66756 standard; Protein; 434 AA.

XX	ABB66756.
XX	
XX	26-MAR-2002 (first entry)
XX	Drosophila melanogaster polypeptide SEQ ID NO 27060.
XX	
XX	Drosophila; developmental biology; cell signalling; insecticide;
XX	pharmaceutical.
XX	
XX	Drosophila melanogaster.
XX	
XX	WO200171042-A2.
XX	
XX	27-SEP-2001.
XX	
XX	23-MAR-2001; 2001WO-US09231.
XX	
XX	23-MAR-2000; 2000US-191637P.
XX	
XX	11-JUL-2000; 2000US-0614150.
XX	
XX	(PEKE) PE CORP NY.
XX	
XX	Venter JC, Adams M, Li PWD, Myers EW;
XX	
XX	WPI; 2001-656860/75.
XX	
XX	N-PSDB; ABL10859.
XX	
XX	PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX	genes from Drosophila and for elucidating cell signalling and cell-cell
XX	interactions -
XX	
XX	Disclosure; SEQ ID NO 27060; 21pp + sequence listing; English.
XX	
XX	The invention relates to an isolated nucleic acid detection reagent
XX	capable of detecting 1000 or more genes from Drosophila. The invention is

cellular in developmental biology and in elucidating cell signaling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (AB116176-AB130511), expressed DNA sequences (AB101840-AB116175) and the encoded proteins (AAB57737-AAB572072).

The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp://wipo.int/pub/published_pat_sequences](http://wipo.int/pub/published_pat_sequences).

Sequence 434 AA;

Query Match	34.8%	Score 1252	DB 22	Length 434
Best Local Similarity	45.6%	Pred. No. 1	7e-52	
Matches 190	Conservative 55	Mismatches 170	Indels 2	Gaps 1

QY	151	KCQCKNGMLCPBIPITGACHCAGAFGMSGCEDECEGGTGINCHQNCOCOMATDCHVTGEC	21.0
QY	151	KCQCKNGMLCPBIPITGACHCAGAFGMSGCEDECEGGTGINCHQNCOCOMATDCHVTGEC	21.0
Db	2	QCQDLNNAVCEPFSFGDCECAKAGTGAACADICPEGFFGAMCSEKCRRENKCHHVSSEC	61
QY	211	RCPPGYTGAFCECIDLCPPEKHGPPQCEQKRCPCQNGVGCCHHVTGEGSCSPGMMGTVCQCPRE	27.0
Db	62	QCAAGFGEPLTCDMHCPRGKHGAQCCQDDPCQNDKQCPETGACMCNPGMTGVDYCANKCPV	12.0
QY	271	GRFERNCSQBEQCHNGSTCDAATQCHCNSRBTGEBRQDECPVGTGTVLCAETQCQVNGG	33.0
Db	122	GSYPCQGESECECKGAPCHHITGQCCPPREYGEREFDECOLNTGFSNMTCDXCANDA	18.0
QY	331	KCYHVSACILCEAEPFAGEBRCBARLCEPGLVIGIKCDKRCPHLENTHSCHPMSGECAKPG	39.0
Db	182	MCDRANGTICNPMTGAKCAKERCCEANKYGLDNCNRCEBDMETHDLCHEPINCOCSTIG	24.0
QY	391	WSGLYCNETSPGFGYGGARQOICSQNGADCDVSATGKTCAPGFKIGDCTPCPLGTGGI	45.0
Db	242	WSSAQCTRPCTFLRYRGPNCLELTCKCKNGKACSPVNGTCLAPGMRGPTCESECEBPFGQ	30.0
QY	451	NCSRRGCCCKNDAYVCSPLVDGSCSTCKAAGHHGYDCSTRPSSGTWGGGCVNLTCOLNGACNLT	51.0
Db	302	DCALRCCQNGAKCEPETHGQCLCTAGAKNKKCDRPPDLNHFQGDCKAKVDCCHNNAACNPQ	36.0
QY	511	DGTGTCAFGNNGEKCCLPQCDGTGLNCAKERCDC--SHAGGCPHTTGHCRLPGWGS	56.5
Db	362	NGSOTCAAGTGGEBCEKCKDTGKFGHNCADKCCQCDPNNSLACATATGRCVCKODMWG	41.8

RESULT 6
AAB66269
ID AAB66269 standard; Protein; 636 AA

AC	AAB66269;	
XX		
DT	05-APR-2001 (first entry)	
XX		
DE	Rat TANGO 272 SEQ ID NO: 20.	
XX		
XX	Membrane associated protein; secreted protein; human; mouse; rat;	
XX	INTERCEPT 340; MANGO 003; MANGO 347; TANGO 272; TANGO 295; TANGO 354;	
XX	TANGO 378; skeletal disorder; cardiovascular disorder; renal disorder	
XX	haemotopoietic disorder; neural disorder; hepatic disorder;	
XX	neoplastic disease.	
OS	Rattus sp.	
XX		
XX		
PN	WO200100673-A1.	
XX		
PD	04-JAN-2001.	
XX		
PF	29-JUN-2000; 2000WO-US18198.	
XX		
PR	30-JUN-1999; 99US-0345464.	
XX		
PA	(MILL-) MILLENNIUM PHARM INC.	
XX		

PI Barnes TM, Fraser CC, Wighton N, Myers P, Busfield SJ, Sharp JD;
 XX WPI: 2001-050128/06.
 DR N-PSDB; AAF27791.
 XX
 PT Isolated secreted or transmembrane proteins are used for diagnosis and
 PT treatment of neoplastic and haematopoietic disorders e.g. T cell
 PT disorders, cancer and tumours -
 XX
 XX Claim 9; Page 238-240; 294pp; English.
 CC The present invention provides the protein and coding sequences for a
 CC number of membrane associated and secreted proteins from human, mouse and
 CC rat. The proteins are designated INTERCEPT 340, MANGO 003, MANGO 347,
 CC TANGO 272, TANGO 295, TANGO 254 and TANGO 378. The proteins are all
 CC involved in signal transduction and the sequences can be used in the
 CC treatment of cardiovascular, renal, hepatic, neural, neoplastic, skeletal
 CC and haematopoietic disorders.
 XX
 XX Sequence 636 AA;

Query Match 33.7%; Score 1215; DB 22; Length 636;
 Best Local Similarity 56.7%; Pred. No. 1.2e-50;
 Matches 185; Conservative 38; Mismatches 103; Indels 0; Gaps 0;

QY 260 MGTVCQGPCEGRRGKNCSEOCNGTCDATGQCSPGYGECRQDECPVGTGYVL 319
 DB 1 MGVCISLPCEGRRGKNCSEOCNGTCDATGQCSPGYGECRQDECPVGTGYVL 319
 QY 320 CAETGOCVNGSKCYHVSACILEAGFAGEEREARLCPGLYGIKDRCPCHLENTSH 379
 DB 61 CAETGOCVNGSKCYHVSACILEAGFAGEEREARLCPGLYGIKDRCPCHLENTSH 379
 QY 380 PMSEGCACKGMSGLYENETCSPEFYGEACQOICSCNGADCDVTKCKCAPBFKIIDC 439
 DB 121 PMSEGCACKGMSGLYENETCSPEFYGEACQOICSCNGADCDVTKCKCAPBFKIIDC 439
 QY 440 STPCPLGTGINSRSGCKNDAYCSPVDSCTCKAGMHGVDCSIRCPSTGTFGNCMLTC 499
 DB 181 ANLCPPMTYGINCSHSCENALICSPVDGTICKBGMQNGNSVPPPGTWGFSNCANSC 240
 QY 500 QCLNGACNLTDTCTCAPRGKCELCPCODGTIYGLNCAERDCSHADGCHPTTGHCRC 559
 DB 241 QCAIEGVCSPTGACTCTCPGWRGVHCOLPCPKGQFEGSCASVCDHSDGDPVHGHCRC 300
 QY 560 LPGMSGVHCDYVCAEGRMGPNCSLPC 585
 DB 301 QAGMGTIRCHLPCPEGFWMGANCNAC 326

RESULT 7
 ABG08031
 ID ABG08031 standard; Protein; 269 AA.

AC ABG08031;

DT 13-FEB-2002 (first entry)

DE Novel human diagnostic protein #8022.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001MO-US08631.

XX 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.
 PA
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI: 2001-639362/73.
 DR N-PSDB; AAF72218.

PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnosis, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX
 XX Claim 20; SEQ ID No 38390; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pcl_sequences.
 XX
 XX Sequence 269 AA;

Query Match 32.5%; Score 1169; DB 22; Length 269;
 Best Local Similarity 85.6%; Pred. No. 1e-48;
 Matches 196; Conservative 3; Mismatches 14; Indels 16; Gaps 3;

QY 338 ACILEAGFAGEEREARLCPGLYGIKDRCPCHLEN--THSCHPMSEGCACKRPMGSLY 395
 DB 5 ALICQLTYA-----C-----ISAQLICPFAMEQOIVACPMHSEGCACKRPMGSLY 50
 QY 396 CNETCSPEFYGEACQOICSCNGADCDVTKCTCAPGFGIDCSTPCPLATYGINCSSR 455
 DB 51 CNETCSPEFYGEACQOICSCNGADCDVTKCTCAPGFGIDCSTPCPLATYGINCSSR 110
 QY 456 CGCKNDAYCSPVDSCTCKAGMHGVDCSIRCPSTGTFGNCMLTCQCLNGACNLTDTCT 515
 DB 111 CGCKNDAYCSPVDSCTCKAGMHGVDCSIRCPSTGTFGNCMLTCQCLNGACNLTDTCT 170
 QY 516 CAPGMRGKCELCPCODGTIYGLNCAERDCSHADGCHPTTGHCRLPMS 564
 DB 171 CAPGMRGKCELCPCODGTIYGLNCAERDCSHADGCHPTTGHCRLPMS 219

RESULT 8
 AAG75479
 ID AAG75479 standard; Protein; 384 AA.

AC AAG75479;

DT 03-SEP-2001 (first entry)

DE Human colon cancer antigen protein SEQ ID NO:6243.

XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
 KW colorectal carcinoma.
 XX

PI Young PE, Ruben SM, Ni J;
XX WPI: 2001-183257/18.
DR N-PSDB: AAD02810.
XX
XX Isolated nucleic acids encoding human attractin-like polypeptides,
PT preventing or treating autoimmune diseases e.g. rheumatoid
PT arthritis, and hyperproliferative disorders -
XX
XX Claim 12; Page 222-223; 235pp; English.
PS
XX The present sequence is a HfICU08 clone human attractin-like
CC protein. This sequence is encoded by a nucleotide sequence located
CC on chromosome 1p36.3. Human attractin-like sequences are useful for
CC preventing, treating or ameliorating a medical condition. Human
CC attractin-like sequences are used as a food additive or preservative
CC to increase or decrease storage capabilities and are also useful for
CC chromosome identification and gene therapy. Human attractin-like
CC sequences and their antibodies, agonists and antagonists are useful in
CC the diagnosis, treatment and prevention of cancer such as ovarian cancer,
CC cancer of breast, adrenal gland, bone, bone marrow, breast, liver, lung,
CC gastrointestinal tract or urogenital system, immune disorders such as
CC Addison's disease, allergies, autoimmune hemolytic anaemia, autoimmune
CC thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis,
CC rheumatoid arthritis and ulcerative colitis, cardiovascular disorders
CC such as myocardial ischaemias, congestive heart failure, pericarditis,
CC wound healing, neurological diseases such as cerebral anoxia,
CC Alzheimer's disease, communicative disorders and epilepsy, and
CC infectious diseases such as viral, bacterial, fungal and parasitic
CC infections.
XX
XX Sequence 474 AA;
SQ

Query Match 31.1%; Score 1119.5; DB 22; Length 474;
Best Local Similarity 44.4%; Pred. No. 3.4e-46;
Matches 195; Conservative 37; Mismatches 196; Indels 11; Gaps 7;

QY 93 QCCP-GEFSEBMCVPHCADKCVHGRCLAP--NFCQCEPMGNGTNCSSACDGDHMRPCT 149
DB 7 QACRAGLY--GDNCRHSL--CQNGCTDDPYSGHCACPEBAGNAGLACEHECLPRDYRACR 62
QY 150 SRQCKNCAICNPITGACHCAAGFERWRCEDRGEGTYGNDCHORQCONCATCDHYTGE 209
DB 63 HSGGCLNGELCDPHGTGRLCPAGMTGDKQSPCLRMGFEGEACARCSCLPAAACHHYTGA 122
QY 210 CRCPGYGARGCEDLCPRGKHGRCQCEORCPQ--KNGVCHNHTGESCSGMMGTVCQGPC 268
DB 123 CRCPGEGFGSGCEQACPPGSGFEGECADQCCQCGENPACHPATGTCTSCAAGYHGPSCQQR 182
QY 269 PEGRGKNCQSOECOCNGTCDAAAGQCHCSPGYTGERCODECPYGTGVCAETCCQCVN 328
DB 183 PPRGRGPCEDLCCGLNGSCDADATGACRCPTGELGTICNLTCPOGRGPRMCTHYCGCGQ 242
QY 329 GAKCYHVSAGCLCEAGFAGERCEARLCPGLYGIKCDRCRCHLENTHSCHPMSECKACK 388
DB 243 GAACDPVYGTCLCPRGARGVRC--RGCPQNRFGVGCETHTSC--RNGGLCHANSNCSG 299
QY 389 PGMGLVYENTCSPGFVEAGAOQICSCONGADSDYTKCTCARGFKIDISTEPPLGT 448
DB 300 LGWGRHCELCAPRGRYGAACHLESCNHNSTCEPATGTCCGPFYQACEHPCPPEFH 359
QY 449 GINTSSRRCNDKAVCSFVDSCTCKAGMHWVDSIRCPSTGFGCM/LTCCCLNGACN 508
DB 360 GAGCGGLCMCHGAPCDISGRCLCPAGFHGFCEGCEPSPFEGCHQRDCDGGAPCD 419
QY 509 TLDTGTCTCAPGMRGKCEL 527
DB 420 PVTGLICLPPGRSGATCNL 438

RESULT 10
ABG03826
ID ABG03826 standard; Protein; 587 AA.

XX
AC ABG03826;
XX
DT 13-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #3817.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN W0200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI: 2001-639362/73.
XX N-PSDB: AAS68013.
XX
XX Claim 20; SEQ ID No 34185; 103pp; English.
PS
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations in
CC and responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.
XX
SQ Sequence 587 AA;

Query Match 27.8%; Score 1002.5; DB 22; Length 587;
Best Local Similarity 31.0%; Pred. No. 1.4e-40;
Matches 214; Conservative 49; Mismatches 178; Indels 249; Gaps 19;

QY 31 PNVCSHMSVSYVQESFPHFDDIYITSCDILNMFCTHRRYSIRAYHGEKTYRR 90
DB 5 PTVLCMAQAGVGV-----RHNIIFWNEKEHG-----HG 34
QY 91 KSQCCPFGFSEGEKCVPHCADKCVHGRCLAPNFCQCEPMGNGTNCSSACDGDHMRPCT 136
DB 35 KSGSC-----HNGASC-----SABEGACHCTPGMTGLCTQRKPHLLASQPLRI 78
QY 137 -----SA 138

CC Alzheimer's disease, Parkinson's disease, stroke, immune disorders,
CC hematopoietic disorders and various dyslipidaemias, metabolic syndrome X
CC and wasting disorders associated with chronic diseases and cancers,
CC cardiac disorders, hypertension, hypercalcaemia, cirrhosis, angiodysplasia,
CC and wound healing, trauma, glomerulonephritis, hyper and hypothyroidism,
CC multiple sclerosis, lung diseases including asthma, Crohn's disease,
CC scleroderma, autoimmune diseases, developmental disorders and neural tube
CC defects. The present sequence is the protein sequence for human MO18b.
CC MO18b is an acetyl LDL receptor-like protein.

XX Sequence 865 AA:

Query Match 22.4%; Score 808; DB 23; Length 865;

Best Local Similarity 33.1%; Pred. No. 3.2e-31;
Matches 156; Conservative 42; Mismatches 164; Indels 110; Gaps 17;

QY 94 CCGFESEEMC-VPHCA--DKCVHGR-CIAVNTQCCERPMGTNCSACDGDHMHPHCT 149
DB 64 CCAGMROGDEGCIANCBENSTCSENEVCVRGECCRGYTGANDTICPQFMPDCK 123
Y 150 SROCKNGALCNPTGACHCAAGFRGWRCEDEGTYGNDCHORCCQNGATCDHVTGE 209
DB 124 ELCSCHPHQCEDVYGCCTCHA--RRW-----GAREHACQCHG-TCFHRSGA 169
QY 210 CCRPGYTGAFCEDLCPPKHPQCEQRPCQNGVCHHTVECCSPGSMGTVCQQPCP 269
DB 170 CRCEPGMWA-----QCASACYCSATSRCDPQTGACICHAQM----- 206
QY 270 EGRFGKNSQECQCHNGTCDATGQCHSPGYTGERCDECPVGYGILCAETCCVNG 329
DB 207 ---WGRSCNNQAC--NSSPEQOSGRQCR-----ER-----TFGARCDRYCCCFRG 249
QY 330 GKCYHVSACILCEAGFAGERCEARLCPEGLYGIKCDKRCPCHELNTHSCHPMSEGCACRP 389
DB 250 -----WGRSCNNQAC--NSSPEQOSGRQCR-----ER-----TFGARCDRYCCCFRG 249
QY 390 GWSGLYCNETGSPGFYGEACQOIC--SCONGADCDSTYKGC-TCAPGFKIDSTPCPLGT 447
DB 263 GYRGKTCRPPCPAGFYGLCRRRCQCKGQPCYVAEGRLCTCEPMWNTKCDPCATGF 322
QY 448 YGINSNRG--GCKNDVAVSPVDGSC--CKAGHGVDCSIRCSGWTGFGCNLTCCCLNG 505
DB 323 YEGGSHRCPDRDGHACHHTVGTCTRCNAWIGDRCEKCSNGTYGDECAFYCADCGSG 382
QY 506 ACNTLDGTCTCAGWRGKCELCPCDGTGYGLNCAERCDSCSHADGCHPTTGHG 557
DB 383 HCDPQSGRLCSPGVHGRPHCVNTCPGLHGAACAQACSC-HEDTCDPVTGAC 433

RESULT 14

XX AAB60393 standard; Protein; 866 AA.

XX AAB60393;

XX 24-APR-2001 (first entry)

XX Human nurse cell receptor B6TNC#10, SEQ ID NO:21.

XX Nurse cell receptor; human; B6TNC#10; chromosome 22q11;

XX DiGeorge's syndrome; drug screening; detection; diagnosis.

XX Homo sapiens.

XX JP2000308492-A.

XX 07-NOV-2000.

XX 23-FEB-2000; 2000JP-0045321.

XX 24-FEB-1999; 99JP-0046604.

XX (SHIO) SHIONOGI & CO LTD.

XX WPI; 2001-161956/17.
DR N-PSDB; AAF27486.

PT New polypeptide which is a human nurse cell receptor is useful for
PT screening assays, especially for DiGeorge's syndrome -

PS Claim 1; Page 26-28; 48pp; Japanese.

CC The invention relates to a human nurse cell receptor protein sequence,
CC designated B6TNC#10 (AAB60393) and to the cDNA clone encoding it
CC (AAF27486). The human B6TNC#10 gene is located on chromosome 22q11,
CC in a region associated with DiGeorge's syndrome (a congenital
CC disorder which is often associated with heart defects, anomalies of
CC the major vessels, oesophageal atresia and abnormalities of facial
CC structures). The human B6TNC#10 gene was identified in a human genomic
CC sequence (Genbank AC005500) via homology with a murine B6TNC#10 cDNA
CC clone (AAF27468) isolated from the murine cell line B6TNC. Appropriate
CC PCR primers (AAF27479-AAF27485) were designed to isolate the human
CC B6TNC#10 cDNA from the SK-LMS-1 cell line. The invention also relates to
CC expression vectors and transformants comprising the human B6TNC#10 cDNA;
CC the recombinant production of human nurse cell receptor B6TNC#10; a
CC monoclonal antibody recognising B6TNC#10; a hydridoma producing such an
CC antibody; a method of screening for a ligand which binds to a human
CC nurse cell receptor; and compounds thus identified. The human B6TNC#10
CC cDNA can be used in the detection and diagnosis of DiGeorge's syndrome.
CC The present sequence represents the human nurse cell receptor
CC encoded by cDNA clone B6TNC#10.

XX Sequence 866 AA:

Query Match 22.4%; Score 808; DB 22; Length 866;

Best Local Similarity 33.1%; Pred. No. 3.2e-31;
Matches 156; Conservative 42; Mismatches 164; Indels 110; Gaps 17;

QY 94 CCGFESEEMC-VPHCA--DKCVHGR-CIAVNTQCCERPMGTNCSACDGDHMHPHCT 149
DB 64 CCAGMROGDEGCIANCBENSTCSENEVCVRGECCRGYTGANDTICPQFMPDCK 123
QY 150 SROCKNGALCNPTGACHCAAGFRGWRCEDEGTYGNDCHORCCQNGATCDHVTGE 209
DB 124 ELCSCHPHQCEDVYGCCTCHA--RRW-----GAREHACQCHG-TCFHRSGA 169
QY 210 CCRPGYTGAFCEDLCPPKHPQCEQRPCQNGVCHHTVECCSPGSMGTVCQQPCP 269
DB 170 CRCEPGMWA-----QCASACYCSATSRCDPQTGACICHAQM----- 206
QY 270 EGRFGKNSQECQCHNGTCDATGQCHSPGYTGERCDECPVGYGILCAETCCVNG 329
DB 207 ---WGRSCNNQAC--NSSPEQOSGRQCR-----ER-----TFGARCDRYCCCFRG 249
QY 330 GKCYHVSACILCEAGFAGERCEARLCPEGLYGIKCDKRCPCHELNTHSCHPMSEGCACRP 389
DB 250 -----WGRSCNNQAC--NSSPEQOSGRQCR-----ER-----TFGARCDRYCCCFRG 249
QY 390 GWSGLYCNETGSPGFYGEACQOIC--SCONGADCDSTYKGC-TCAPGFKIDSTPCPLGT 447
DB 263 GYRGKTCRPPCPAGFYGLCRRRCQCKGQPCYVAEGRLCTCEPMWNTKCDPCATGF 322
QY 448 YGINSNRG--GCKNDVAVSPVDGSC--CKAGHGVDCSIRCSGWTGFGCNLTCCCLNG 505
DB 323 YEGGSHRCPDRDGHACHHTVGTCTRCNAWIGDRCEKCSNGTYGDECAFYCADCGSG 382
QY 506 ACNTLDGTCTCAGWRGKCELCPCDGTGYGLNCAERCDSCSHADGCHPTTGHG 557
DB 383 HCDPQSGRLCSPGVHGRPHCVNTCPGLHGAACAQACSC-HEDTCDPVTGAC 433

RESULT 15

XX AAB60394 standard; Protein; 866 AA.

XX AAB60394;

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4.P5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 9, 2003, 11:54:50 ; Search time 63.1495 Seconds

(without alignments)
853.959 Million cell updates/sec

Title: US-10-092-390-4

Perfect score: 3601
Sequence: 1 MVISLNSCLSTICLLCHMT.....HDSVCAEGRWGPNCLSPCY 586

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 349150 seqs, 92025710 residues

Total number of hits satisfying chosen parameters: 349150

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: Published_Applications_AA:*

- 1: /cgn2_6/ptodata/2/pubppaa/US08_NEW_PUB pep:*
- 2: /cgn2_6/ptodata/2/pubppaa/PCF_NEW_PUB pep:*
- 3: /cgn2_6/ptodata/2/pubppaa/US06_NEW_PUB pep:*
- 4: /cgn2_6/ptodata/2/pubppaa/US06_PUBCOMB pep:*
- 5: /cgn2_6/ptodata/2/pubppaa/US07_NEW_PUB pep:*
- 6: /cgn2_6/ptodata/2/pubppaa/US07_PUBCOMB pep:*
- 7: /cgn2_6/ptodata/2/pubppaa/PCUS_PUBCOMB pep:*
- 8: /cgn2_6/ptodata/2/pubppaa/US08_PUBCOMB pep:*
- 9: /cgn2_6/ptodata/2/pubppaa/US09_NEW_PUB pep:*
- 10: /cgn2_6/ptodata/2/pubppaa/US09_PUBCOMB pep:*
- 11: /cgn2_6/ptodata/2/pubppaa/US10_NEW_PUB pep:*
- 12: /cgn2_6/ptodata/2/pubppaa/US10_PUBCOMB pep:*
- 13: /cgn2_6/ptodata/2/pubppaa/US60_NEW_PUB pep:*
- 14: /cgn2_6/ptodata/2/pubppaa/US60_PUBCOMB pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3601	100.0	586	9 US-10-092-390-4	Sequence 4, Appl1
2	3601	100.0	1140	9 US-10-092-390-2	Sequence 2, Appl1
3	1667.5	46.3	1050	9 US-09-796-753-114	Sequence 114, App
4	1215	33.7	636	9 US-09-796-753-100	Sequence 100, App
5	1215	33.7	636	9 US-09-796-753-124	Sequence 124, App
6	1119.5	31.1	474	9 US-10-084-994-11	Sequence 11, Appl1
7	843.5	23.4	830	9 US-09-870-759-134	Sequence 134, App
8	843.5	23.4	830	9 US-09-870-759-140	Sequence 140, App
9	843.5	23.4	830	9 US-09-842-758-57	Sequence 57, Appl1
10	808	22.4	865	9 US-09-842-758-20	Sequence 20, Appl1
11	783.5	21.6	934	9 US-09-842-758-18	Sequence 18, Appl1
12	779	21.6	299	9 US-09-866-050A-458	Sequence 458, App
13	779	21.6	299	9 US-09-866-050A-192	Sequence 192, App
14	779	21.6	299	9 US-09-866-050A-332	Sequence 332, App
15	721.5	20.0	310	9 US-10-084-994-12	Sequence 12, Appl1
16	664.5	18.5	572	9 US-09-900-449A-7	Sequence 7, Appl1
17	662	18.4	2444	10 US-09-944-849-2	Sequence 2, Appl1
18	656.5	18.2	601	9 US-09-900-449A-5	Sequence 5, Appl1
19	656	18.2	241	9 US-10-084-994-8	Sequence 8, Appl1

20	645.5	17.9	2167	10 US-09-778-927A-61	Sequence 61, Appl1
21	645	17.9	534	9 US-10-125-459-6	Sequence 6, Appl1
22	645	17.9	534	9 US-10-067-761-14	Sequence 14, Appl1
23	645	17.9	534	10 US-09-804-156-14	Sequence 14, Appl1
24	645	17.9	534	10 US-09-946-633-6	Sequence 6, Appl1
25	644.5	17.9	2201	12 US-10-100-912-2	Sequence 2, Appl1
26	638	17.7	2743	9 US-10-037-182-36	Sequence 36, Appl1
27	638	17.7	3695	9 US-10-037-182-2	Sequence 2, Appl1
28	634	17.6	1055	9 US-10-219-248-2	Sequence 2, Appl1
29	634	17.6	1055	9 US-10-219-247-2	Sequence 2, Appl1
30	634	17.6	1055	10 US-09-855-722-2	Sequence 2, Appl1
31	633	17.6	1212	9 US-10-219-248-3	Sequence 3, Appl1
32	633	17.6	1212	9 US-10-219-247-3	Sequence 3, Appl1
33	633	17.6	1212	10 US-09-855-722-3	Sequence 3, Appl1
34	633	17.5	1238	10 US-09-844-849-4	Sequence 4, Appl1
35	630.5	17.5	639	9 US-09-900-449A-4	Sequence 4, Appl1
36	629	17.5	1036	10 US-09-995-593A-6	Sequence 6, Appl1
37	629	17.5	1187	10 US-09-995-593A-7	Sequence 7, Appl1
38	629	17.5	1208	9 US-10-213-329-1	Sequence 1, Appl1
39	629	17.5	1218	10 US-09-995-593A-11	Sequence 11, Appl1
40	629	17.5	1218	10 US-09-944-849-3	Sequence 3, Appl1
41	622	17.3	1238	9 US-10-219-248-5	Sequence 5, Appl1
42	622	17.3	1238	9 US-10-219-247-5	Sequence 5, Appl1
43	622	17.3	1238	10 US-09-855-722-5	Sequence 5, Appl1
44	619	17.2	1218	9 US-10-219-248-7	Sequence 7, Appl1
45	619	17.2	1218	9 US-10-219-247-7	Sequence 7, Appl1

ALIGNMENTS

RESULT 1
US-10-092-390-4
; Sequence 4, Application US/10092390
; Publication No. US20030013865A1
; GENERAL INFORMATION:
; APPLICANT: Yu, Xuanchuan
; TITLE OR INVENTION: No. US20030013865A1 Human EGF-Family Proteins and Polynucle
; FILE REFERENCE: LEX-0317-USA
; CURRENT APPLICATION NUMBER: US/10/092,390
; PRIOR FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: US 60/275,013
; PRIORITY FILING DATE: 2001-03-12
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 586
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-092-390-4

Query Match 100.0%; Score 3601; DB 9; Length 586;
Best Local Similarity 100.0%; Pred. No. 3e-175;
Matches 586; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVISLNSCLSTICLLCHMTGASPLNEDNVSCHMSYVYQESYPRHFDIYTSC 60
DB 1 MVISLNSCLSTICLLCHMTGASPLNEDNVSCHMSYVYQESYPRHFDIYTSC 60
QY 1 TDILNFKCTHRHRSYRAVHGEKTYRRRSQCCPGFYESEGCVPFHCAADKCHGRCIA 120
DB 1 TDILNFKCTHRHRSYRAVHGEKTYRRRSQCCPGFYESEGCVPFHCAADKCHGRCIA 120
QY 121 PNTQCEPFGMGWTCSSACDDHMGPHCTSRQCKNKAALCPITGACCAAGFGWRCED 180
DB 121 PNTQCEPFGMGWTCSSACDDHMGPHCTSRQCKNKAALCPITGACCAAGFGWRCED 180
QY 181 RCEGTGTGNDCHOCQONATGCHVYGECCPGYTGATCEDCPRGKRGPOEQRCP 240
DB 181 RCEGTGTGNDCHOCQONATGCHVYGECCPGYTGATCEDCPRGKRGPOEQRCP 240
QY 241 QNGSVCHHYTGECSCPSGMWGTVCGPCPEGRFKNCSQECQCHNGCTCAATGQCCHCSP 300

```

Db 241 QNGGVCCHHTGECSCPSGMMGTVCQPCPEGRFGKNCSECCCHNGGTDAATGQCHCSP
Oy 301 GTTGERCODECPVGTGYVLCATCCOCVNGKCYHVSAGACLCAGFAGERCARLCEGLY 360
Db 301 GTTGERCODECPVGTGYVLCATCCOCVNGKCYHVSAGACLCAGFAGERCARLCEGLY 360
Oy 361 GIKDKRCPCHELENTSHCHPMSEGCACRPGMSGLYCNETCSPGFYGEACQIICSCONGAD 420
Db 361 GIKDKRCPCHELENTSHCHPMSEGCACRPGMSGLYCNETCSPGFYGEACQIICSCONGAD 420
Oy 421 CDSVGTGKTCACGAFGIDSTPCPLGTGGINSSRCGCKNDVAVCSVPDSCCTCKAGMHGV 480
Db 421 CDSVGTGKTCACGAFGIDSTPCPLGTGGINSSRCGCKNDVAVCSVPDSCCTCKAGMHGV 480
Oy 481 DCSINCPSTGTFGCLNTCCCLNGAGACNTLDGTCTCAPGMRGKELPCQDGTGYNCAE 540
Db 481 DCSINCPSTGTFGCLNTCCCLNGAGACNTLDGTCTCAPGMRGKELPCQDGTGYNCAE 540
Oy 541 RDCSHADGCHPTTGHCRCLPGMSGVHCDVCAEGRMGFNCSLPCY 586
Db 541 RDCSHADGCHPTTGHCRCLPGMSGVHCDVCAEGRMGFNCSLPCY 586

```

RESULT 2

```

US-10-092-390-2
; Sequence 2, Application US/10092390
; Publication No. US20030013865A1
; GENERAL INFORMATION:
; APPLICANT: Yu, Xuanchuan
; TITLE OF INVENTION: No. US20030013865A1 Human EGF-Family Proteins and Polynucleotid
; FILE REFERENCE: Lex-0317-USA
; CURRENT APPLICATION NUMBER: US/10/092,390
; CURRENT FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: US 60/275,013
; PRIOR FILING DATE: 2001-03-12
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 2
; LENGTH: 1140
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-092-390-2

```

Query Match

```

Best Local Similarity 100.0%; Score 3601; DB 9; Length 1140;
Pred. No. 4,8e-175;
Matches 586; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Oy 1 WTISLNSCLSTFICLLCHWIGTASPLNLEDPVNCSHWESYSTVQESYHPFDQIYYTSC 60
Db 1 WTISLNSCLSTFICLLCHWIGTASPLNLEDPVNCSHWESYSTVQESYHPFDQIYYTSC 60
Oy 61 TDLNMFKCTRRVSRVRYRGEKTMRYRKSOCCPGFESGEMCVPHCADCVHGRCTA 120
Db 61 TDLNMFKCTRRVSRVRYRGEKTMRYRKSOCCPGFESGEMCVPHCADCVHGRCTA 120
Oy 121 PNTCCPEPMGWTGNCSSACDGDHMGPHCTSRQCKNGALCNITGACHCAAGFRNGRCD 180
Db 121 PNTCCPEPMGWTGNCSSACDGDHMGPHCTSRQCKNGALCNITGACHCAAGFRNGRCD 180
Oy 121 PNTCCPEPMGWTGNCSSACDGDHMGPHCTSRQCKNGALCNITGACHCAAGFRNGRCD 180
Db 121 PNTCCPEPMGWTGNCSSACDGDHMGPHCTSRQCKNGALCNITGACHCAAGFRNGRCD 180
Oy 181 RCEQGTGNDCHQRCQNGATCDHYTGRCRCPGTGTAFCEDLCPRGHNGQCEQRCPC 240
Db 181 RCEQGTGNDCHQRCQNGATCDHYTGRCRCPGTGTAFCEDLCPRGHNGQCEQRCPC 240
Oy 241 QNGGVCCHHTGECSCPSGMMGTVCQPCPEGRFGKNCSECCCHNGGTDAATGQCHCSP 300
Db 241 QNGGVCCHHTGECSCPSGMMGTVCQPCPEGRFGKNCSECCCHNGGTDAATGQCHCSP 300
Oy 301 GTTGERCODECPVGTGYVLCATCCOCVNGKCYHVSAGACLCAGFAGERCARLCEGLY 360
Db 301 GTTGERCODECPVGTGYVLCATCCOCVNGKCYHVSAGACLCAGFAGERCARLCEGLY 360

```

```

Oy 361 GIKDKRCPCHELENTSHCHPMSEGCACRPGMSGLYCNETCSPGFYGEACQIICSCONGAD 420
Db 361 GIKDKRCPCHELENTSHCHPMSEGCACRPGMSGLYCNETCSPGFYGEACQIICSCONGAD 420
Oy 421 CDSVGTGKTCACGAFGIDSTPCPLGTGGINSSRCGCKNDVAVCSVPDSCCTCKAGMHGV 480
Db 421 CDSVGTGKTCACGAFGIDSTPCPLGTGGINSSRCGCKNDVAVCSVPDSCCTCKAGMHGV 480
Oy 481 DCSINCPSTGTFGCLNTCCCLNGAGACNTLDGTCTCAPGMRGKELPCQDGTGYNCAE 540
Db 481 DCSINCPSTGTFGCLNTCCCLNGAGACNTLDGTCTCAPGMRGKELPCQDGTGYNCAE 540
Oy 541 RDCSHADGCHPTTGHCRCLPGMSGVHCDVCAEGRMGFNCSLPCY 586
Db 541 RDCSHADGCHPTTGHCRCLPGMSGVHCDVCAEGRMGFNCSLPCY 586

```

RESULT 3

```

US-09-796-753-114
; Sequence 114, Application US/09796753
; Publication No. US20030027998A1
; GENERAL INFORMATION:
; APPLICANT: McCarthy, Sean A.
; TITLE OF INVENTION: SECRETED PROTEINS AND USES THEREOF
; FILE REFERENCE: 7853-227-999
; CURRENT APPLICATION NUMBER: US/09/796,753
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 09/183,175
; PRIOR FILING DATE: 1998-10-30
; PRIOR APPLICATION NUMBER: 09/223,094
; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 09/223,546
; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 09/224,246
; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 09/259,388
; PRIOR FILING DATE: 1999-02-26
; PRIOR APPLICATION NUMBER: 60/122,458
; PRIOR FILING DATE: 1999-03-01
; PRIOR APPLICATION NUMBER: 09/312,359
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 09/336,536
; PRIOR FILING DATE: 1999-06-18
; PRIOR APPLICATION NUMBER: 09/342,687
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: 09/345,464
; PRIOR FILING DATE: 1999-06-30
; PRIOR APPLICATION NUMBER: 09/365,164
; PRIOR FILING DATE: 1999-07-30
; PRIOR APPLICATION NUMBER: 09/399,723
; PRIOR FILING DATE: 1999-09-20
; PRIOR APPLICATION NUMBER: 09/409,634
; PRIOR FILING DATE: 1999-09-30
; PRIOR APPLICATION NUMBER: 09/471,179
; PRIOR FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: 08/474,071
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 09/474,072
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 09/514,010
; PRIOR FILING DATE: 2000-02-25
; PRIOR APPLICATION NUMBER: 08/516,745
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 09/572,002
; PRIOR FILING DATE: 2000-05-14
; PRIOR APPLICATION NUMBER: 09/597,993
; PRIOR FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: 09/599,596
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 09/630,334
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: 09/606,565
; PRIOR FILING DATE: 2000-06-29

```

PRIOR APPLICATION NUMBER: 09/606,317
 PRIOR FILING DATE: 2000-06-29
 PRIOR APPLICATION NUMBER: 09/665,666
 PRIOR FILING DATE: 2000-09-20
 PRIOR APPLICATION NUMBER: 09/677,751
 PRIOR FILING DATE: 2000-09-30
 NUMBER OF SEQ ID NOS: 162
 SEQ ID NO 114
 LENGTH: 1050
 TYPE: prt
 ORGANISM: Homo sapiens
 US-09-796-753-114

Query Match 46.3%; Score 1667.5; DB 9; Length 1050;
 Best Local Similarity 45.0%; Pred. No. 2.1e-77;
 Matches 284; Conservative 61; Mismatches 181; Indels 105; Gaps 8;

14 LLLCHWIGTASPLNLEDPVNCSHMESVSYVDSYPHPDQIYYTSCDILNW-----66
 9 LLLAVGLRAGLTNPSPDPTCSFMESFTTTTKSHSRPFLSEPC--RPMSEPHTCR 66
 67 -----FKTRHVSRY-----TAY 80
 67 SPQTKRLASRDSFCMVCVAGVOWRDSALOPOTGNALSMRPPRVLSGAPSLASPGH 126
 81 RHGEKTYRRKSQCCPGEYSESGEMCVPHCADKCVHGRCIAPNTCOCEPBGNGSSA-- 138
 127 TVYVKTDHNRLOCCGHEFESRGFCVPLAOECVHGRCAVAPNOCQVPCGRGDDSSAPN 186
 139 ----CDGDMHGRHCTSRCCCKNGKALCNPTGACHCAAGFRGWRCEDERCDEGTGANCOR 194
 187 CLPCTPRGYPRACQPRCC--HGAPCDPQTGAFCFAERTGPRCDVSCSGT-----227
 195 CQONATDHTGECRCRPGYTGAFCEDLCPRGKGRPOCEORCPQNGVCHNHTGEGS 254
 238 -----SGEFC-----PSTH-----PCQNGGVFQTPQGS 262
 255 CFSGMNGVCOGRCPEGRFGKNCQOCNGTCDATGQCHSCSYTGERCQDCPCPVG 314
 263 CRRGMNGTICSLPCRFEGFNGNSQRCRCHNGSLDRFTGQCCACAGTYDRKRECPVG 322
 315 TVGVLCAEFTQCVNGSKCYHVSAGLCEAGFAGERCEANLCREGLYIKDKRCPCHLN 374
 323 RFQDCAETTCDAKPDARCPANAGCLCEHGFETGDRCTDRICRPGFYGLSCQAPCTCDREH 382
 375 THSCHMSECCACKRPMGSLYCHETSPGFYGAACQIQISCONGACCDSTYTGKTAARPG 434
 383 SLSCHPMNECCSLPFWAGLHCNESCRODTHGPGCQEHCLILHGVCAOTSGLCQCAPGY 442
 435 KGLDSTPCPLATGYGINSRCRCCKNDAYSPVDSCTCKAGHNGVDSTIRCPSTWFG 494
 443 TGFHCASLERPDTYGVNCAKRCENAIACSPIDGECVCKEGRQNCVPRCPGTWGS 502
 495 CNITCOCLNGACNTLDGTCTCAPGRGKRCCLPCQDGYTYLNCARCDSCSHADCHPTT 554
 503 CNAKSCCAIBANVCSPTGTACTCPRGHNGANCOLPCPKQGFEGECACRCDHSDGDPVN 562
 555 GHORCLPGRWSGVHCDVSCAEGRWGRNCSLPC 585
 563 GRCQCAQWAGMARCLHSCPEGLMGNVCSNTG 593

RESULT 4
 US-09-796-753-100
 Sequence 100, Application US/09796753
 Publication No. US20030027998A1
 GENERAL INFORMATION:
 APPLICANT: McCarthy, Sean A.
 TITLE OF INVENTION: SECRETED PROTEINS AND USES THEREOF
 FILE REFERENCE: 7853-227-999
 CURRENT APPLICATION NUMBER: US/09/796,753
 PRIOR FILING DATE: 2001-03-01
 PRIOR APPLICATION NUMBER: 09/183,175

PRIOR FILING DATE: 1998-10-30
 PRIOR APPLICATION NUMBER: 09/223,094
 PRIOR FILING DATE: 1998-12-30,546
 PRIOR APPLICATION NUMBER: 09/223,546
 PRIOR FILING DATE: 1998-12-30
 PRIOR APPLICATION NUMBER: 09/224,246
 PRIOR FILING DATE: 1998-12-30
 PRIOR APPLICATION NUMBER: 09/259,388
 PRIOR FILING DATE: 1999-02-26
 PRIOR APPLICATION NUMBER: 60/122,458
 PRIOR FILING DATE: 1999-03-01
 PRIOR APPLICATION NUMBER: 09/312,359
 PRIOR FILING DATE: 1999-05-14
 PRIOR APPLICATION NUMBER: 09/336,536
 PRIOR FILING DATE: 1999-06-18
 PRIOR APPLICATION NUMBER: 09/342,687
 PRIOR FILING DATE: 1999-06-29
 PRIOR APPLICATION NUMBER: 09/345,464
 PRIOR FILING DATE: 1999-06-30
 PRIOR APPLICATION NUMBER: 09/365,164
 PRIOR FILING DATE: 1999-07-30
 PRIOR APPLICATION NUMBER: 09/399,723
 PRIOR FILING DATE: 1999-09-20
 PRIOR APPLICATION NUMBER: 09/409,634
 PRIOR FILING DATE: 1999-09-30
 PRIOR APPLICATION NUMBER: 09/471,179
 PRIOR FILING DATE: 1999-12-23
 PRIOR APPLICATION NUMBER: 09/474,071
 PRIOR FILING DATE: 1999-12-29
 PRIOR APPLICATION NUMBER: 09/474,072
 PRIOR FILING DATE: 1999-12-29
 PRIOR APPLICATION NUMBER: 09/514,010
 PRIOR FILING DATE: 2000-02-25
 PRIOR APPLICATION NUMBER: 09/516,745
 PRIOR FILING DATE: 2000-03-01
 PRIOR APPLICATION NUMBER: 09/572,002
 PRIOR FILING DATE: 2000-05-14
 PRIOR APPLICATION NUMBER: 09/597,993
 PRIOR FILING DATE: 2000-06-19
 PRIOR APPLICATION NUMBER: 09/599,596
 PRIOR FILING DATE: 2000-06-22
 PRIOR APPLICATION NUMBER: 09/630,334
 PRIOR FILING DATE: 2000-07-31
 PRIOR APPLICATION NUMBER: 09/606,565
 PRIOR FILING DATE: 2000-06-29
 PRIOR APPLICATION NUMBER: 09/606,317
 PRIOR FILING DATE: 2000-06-29
 PRIOR APPLICATION NUMBER: 09/665,666
 PRIOR FILING DATE: 2000-09-20
 PRIOR APPLICATION NUMBER: 09/677,751
 PRIOR FILING DATE: 2000-09-30
 NUMBER OF SEQ ID NOS: 162
 SEQ ID NO 100
 LENGTH: 636
 TYPE: prt
 ORGANISM: Rattus sp.
 US-09-796-753-100

Query Match 33.7%; Score 1215; DB 9; Length 636;
 Best Local Similarity 56.7%; Pred. No. 1.1e-54;
 Matches 185; Conservative 38; Mismatches 103; Indels 0; Gaps 0;

260 MGVTCGQPCPEGRFGKNCQECQCHNGTCDATGQCHSCSPYTGRCODECPVGYGV 319
 1 MGVTCGQPCPEGRFGKNCQECQCHNGTCDATGQCHSCSPYTGRCODECPVGYGV 60
 320 CAETCCVNGGKCYHVSAGLCEAGFAGERCEANLCREGLYIKDKRCPCHLNTWHSCH 379
 61 CAETCCVNGGKCYHVSAGLCEAGFAGERCEANLCREGLYIKDKRCPCHLNTWHSCH 120
 380 PMSGCACRPMGWSGLYCNCTSPGFYGAACQIQISCONGACCDSTYTGKTAARPGKIDC 439
 121 PMHGECCQPRWAGLHCNESCRODTHGPGCQEHCLILHGVCAOTSGLCQCAPGYTGPC 180

```

OY 440 STPCPLGTGTCSSRCCKNDVAVSPYDGSCTCKAGHGVDCSIRCPSTGWGECNLTJC 499
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: 09/677,751
; PRIOR FILING DATE: 2000-09-30
; NUMBER OF SEQ ID NOS: 162
; SEQ ID NO 124
; LENGTH: 636
; TYPE: PRF
; ORGANISM: Rattus sp.
; US-09-796-753-124
Db 301 QAGMGTGRCHLPCEGFGWGANCSNAC 326

OY 500 QCLNGAGCNTIDGTCTCAPRGKEKCELPDGTGTLNCARCCDSHADGCHPTTGHCRC 559
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: 09/677,751
; PRIOR FILING DATE: 2000-09-30
; NUMBER OF SEQ ID NOS: 162
; SEQ ID NO 124
; LENGTH: 636
; TYPE: PRF
; ORGANISM: Rattus sp.
; US-09-796-753-124
Db 241 QCAHEGVCSPTGTACTCTPGMRGVHCOLPCPKGFGEGCASVCDCHSDGCDPVHGHCR 300

OY 560 LFGWSGVHDSVCAEGRMGPNCSLPC 585
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: 09/677,751
; PRIOR FILING DATE: 2000-09-30
; NUMBER OF SEQ ID NOS: 162
; SEQ ID NO 124
; LENGTH: 636
; TYPE: PRF
; ORGANISM: Rattus sp.
; US-09-796-753-124
Db 301 QAGMGTGRCHLPCEGFGWGANCSNAC 326

RESULT 5
; Sequence 124, Application US/09796753
; Publication No. US20030027998A1
; GENERAL INFORMATION:
; APPLICANT: McCarthy, Sean A.
; TITLE OF INVENTION: SECRETED PROTEINS AND USES THEREOF
; FILE REFERENCE: 7853-227-999
; CURRENT APPLICATION NUMBER: US/09/796,753
; PRIOR FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 09/183,175
; PRIOR FILING DATE: 1998-10-30
; PRIOR APPLICATION NUMBER: 09/223,094
; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 09/223,546
; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 09/224,246
; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 09/259,388
; PRIOR FILING DATE: 1999-02-26
; PRIOR APPLICATION NUMBER: 60/122,458
; PRIOR FILING DATE: 1999-03-01
; PRIOR APPLICATION NUMBER: 09/312,359
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 09/336,536
; PRIOR FILING DATE: 1999-06-18
; PRIOR APPLICATION NUMBER: 09/342,687
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: 09/345,464
; PRIOR FILING DATE: 1999-06-30
; PRIOR APPLICATION NUMBER: 09/365,164
; PRIOR FILING DATE: 1999-07-30
; PRIOR APPLICATION NUMBER: 09/399,723
; PRIOR FILING DATE: 1999-09-20
; PRIOR APPLICATION NUMBER: 09/409,634
; PRIOR FILING DATE: 1999-09-30
; PRIOR APPLICATION NUMBER: 09/471,179
; PRIOR FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: 09/474,071
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 09/474,072
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 09/514,010
; PRIOR FILING DATE: 2000-02-25
; PRIOR APPLICATION NUMBER: 09/516,745
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 09/572,002
; PRIOR FILING DATE: 2000-05-14
; PRIOR APPLICATION NUMBER: 09/597,993
; PRIOR FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: 09/599,596
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 09/630,334
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: 09/606,565
; PRIOR FILING DATE: 2000-06-29
; PRIOR APPLICATION NUMBER: 09/606,317
; PRIOR FILING DATE: 2000-06-29
; PRIOR APPLICATION NUMBER: 09/665,666

```

```

; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: 09/677,751
; PRIOR FILING DATE: 2000-09-30
; NUMBER OF SEQ ID NOS: 162
; SEQ ID NO 124
; LENGTH: 636
; TYPE: PRF
; ORGANISM: Rattus sp.
; US-09-796-753-124
Query Match 33.7% Score 1215; DB 9; Length 636;
Best Local Similarity 56.7%; Pred. No. 1,1e-54;
Matches 185; Conservative 38; Mismatches 103; Indels 0; Gaps 0;

OY 260 MGTVCQPCPEGRGKNCQDCCQCHNGTCDATGCHSCSPYTGRCODECPVGTGYVL 319
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: 09/677,751
; PRIOR FILING DATE: 2000-09-30
; NUMBER OF SEQ ID NOS: 162
; SEQ ID NO 124
; LENGTH: 636
; TYPE: PRF
; ORGANISM: Rattus sp.
; US-09-796-753-124
Db 1 MGVCISLPCEPBGFGHNPCTGCRCHNGSLDPRFGQCCACAGYIGDRCRECPVGRGQD 60

OY 320 CAETCCQVNGCKCYHVSACLCCEGFAEGERCEANLCPBGLYIGCDKRCPCHEINTSHCH 379
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: 09/677,751
; PRIOR FILING DATE: 2000-09-30
; NUMBER OF SEQ ID NOS: 162
; SEQ ID NO 124
; LENGTH: 636
; TYPE: PRF
; ORGANISM: Rattus sp.
; US-09-796-753-124
Db 61 CAETCDAPGARCPNAGACICEHGTGDRCTERLCPDGRYGLSCDPCTCDPHEISLCH 120

OY 380 PMSECAKCPMGSGLYCNENCSPGFYGACQIQSCONGADCDSDVTGKTCAPBFGKIDC 439
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: 09/677,751
; PRIOR FILING DATE: 2000-09-30
; NUMBER OF SEQ ID NOS: 162
; SEQ ID NO 124
; LENGTH: 636
; TYPE: PRF
; ORGANISM: Rattus sp.
; US-09-796-753-124
Db 121 PMHECCSQPQWAGLHCNESCPODTHGACQCHCLCHGVCLADSGLCRCAPBGTGPHC 180

OY 440 STPCPLGTGTCSSRCCKNDVAVSPYDGSCTCKAGHGVDCSIRCPSTGWGECNLTJC 499
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: 09/677,751
; PRIOR FILING DATE: 2000-09-30
; NUMBER OF SEQ ID NOS: 162
; SEQ ID NO 124
; LENGTH: 636
; TYPE: PRF
; ORGANISM: Rattus sp.
; US-09-796-753-124
Db 181 ANLCPMTYTGTCSSRCCKNDVAVSPYDGSCTCKAGHGVDCSIRCPSTGWGECNLTJC 240

OY 500 QCLNGAGCNTIDGTCTCAPRGKEKCELPDGTGTLNCARCCDSHADGCHPTTGHCRC 559
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: 09/677,751
; PRIOR FILING DATE: 2000-09-30
; NUMBER OF SEQ ID NOS: 162
; SEQ ID NO 124
; LENGTH: 636
; TYPE: PRF
; ORGANISM: Rattus sp.
; US-09-796-753-124
Db 241 QCAHEGVCSPTGTACTCTPGMRGVHCOLPCPKGFGEGCASVCDCHSDGCDPVHGHCR 300

OY 560 LFGWSGVHDSVCAEGRMGPNCSLPC 585
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: 09/677,751
; PRIOR FILING DATE: 2000-09-30
; NUMBER OF SEQ ID NOS: 162
; SEQ ID NO 124
; LENGTH: 636
; TYPE: PRF
; ORGANISM: Rattus sp.
; US-09-796-753-124
Db 301 QAGMGTGRCHLPCEGFGWGANCSNAC 326

RESULT 6
; Sequence 11, Application US/10084994
; Publication No. US20030023070A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: Attractin-like Polynucleotides, Polypeptides, and Antibodies
; FILE REFERENCE: PTO11P1C1
; CURRENT APPLICATION NUMBER: US/10/084,994
; PRIOR FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/790,621
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: PCT/US00/23663
; PRIOR FILING DATE: 2000-08-29
; PRIOR APPLICATION NUMBER: US 60/151,348
; PRIOR FILING DATE: 1999-08-30
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: Patentln Ver. 3.1
; SEQ ID NO 11
; LENGTH: 474
; TYPE: PRF
; ORGANISM: Homo sapiens
; US-10-084-994-11
Query Match 31.1% Score 1119.5; DB 9; Length 474;
Best Local Similarity 44.4%; Pred. No. 5.8e-50;
Matches 195; Conservative 37; Mismatches 196; Indels 11; Gaps 7;

OY 93 QCCP-GEYSEGEKCVPHCADKCVHGRCIAP--NTCCPEPMGSGTSSACDGDHMBHCT 149
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: 09/677,751
; PRIOR FILING DATE: 2000-09-30
; NUMBER OF SEQ ID NOS: 162
; SEQ ID NO 124
; LENGTH: 636
; TYPE: PRF
; ORGANISM: Rattus sp.
; US-09-796-753-124
Db 7 QACPAQIY--GDNCRHSLT--CONGTCDPVSHGACPEPMAGIACKEKELPRDVRAGR 62

OY 150 SROCKNGALCNPTTGACHCAAGFRWRKCDREOGIGYNDCHQRCCQAGATCDHTTGE 209
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: 09/677,751
; PRIOR FILING DATE: 2000-09-30
; NUMBER OF SEQ ID NOS: 162
; SEQ ID NO 124
; LENGTH: 636
; TYPE: PRF
; ORGANISM: Rattus sp.
; US-09-796-753-124

```


Db 63 HSGGCLNGSLDPRHTGRICLPAGWTBDRKQSPCLRWGFEAGCAORCSCLPGACHNHYGA 122

QY 210 CRCPRTYGTAGCEBCLCPRGKRGPCQEDRCPCQ -NGVCHNHTGSCSPSGMGVTCQPC 268

Db 123 CRCPRTYGTAGCEBCLCPRGKRGPCQEDRCPCQ -NGVCHNHTGSCSPSGMGVTCQPC 182

QY 269 PEGFRGNKCSQECQCHNGGTCDAATGQCHSPGYTGERCODECPVGTGYVLAETCOCVN 328

Db 183 PEGFRGNKCSQECQCHNGGTCDAATGQCHSPGYTGERCODECPVGTGYVLAETCOCVN 242

QY 329 GGRKYHVSAGLCEAGFAGERCEARLCPGLYGIKCDKRCPCHELNTHSCHPMSECAK 388

Db 243 GAACDPVTGTCCLPPRGAGVRC -RGCPRNFRGVGCEHTSC -RNGGLCHANSQSCG 299

QY 389 PGNGLICNENCSRGFTGEACQOICSCONAGDCSVTCKTCAAGFGIDCSTPCPLGT 448

Db 300 LGWGRHCELCAPRGRRGAACHLECSCHNSTCEPATGTCCGFTGQACENHCPGFGH 359

QY 449 GINCSRCGCKNDVCSFVDSCTCKAGMHGVDCSIRCPSTWGFGLTQCLNGACN 508

Db 360 GAGCGGLCKWCHGARPCDIPISRCICPAGFHGFCERCEGSRFEGCHQRCDGAPCD 419

QY 509 TLDGTCTCAPGMRKCEL 527

Db 420 PVTGLCLPPRGSGATCNL 438

RESULT 7

US-09-870-759-134

; Sequence 134, Application US/09870759

; Patent No. US20020177551A1

; GENERAL INFORMATION:

; APPLICANT: TERMAN, David S

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT OF NEOPLASTIC DISEASE

; FILE REFERENCE: 870759

; CURRENT APPLICATION NUMBER: US/09/870,759

; CURRENT FILING DATE: 2002-01-14

; PRIOR APPLICATION NUMBER: US 60/208,128

; PRIOR FILING DATE: 2000-05-30

; NUMBER OF SEQ ID NOS: 166

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 134

; LENGTH: 830

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-870-759-134

Query Match 23.4%; Score 843.5; DB 9; Length 830;

Best Local Similarity 36.2%; Pred. No. 7.4e-36;

Matches 158; Conservative 51; Mismatches 164; Indels 63; Gaps 15;

Db 285 EPGWNGTQCCQPCLLPBTGFESECEQCPHCRHGEACEPDTGHCQRCDPGLMRCEDEPCPT 344

QY 446 GTYINCSRRGCKNDVCSFVDSCTCKAGMHGVDCSIRCPSTWGFGLTQCLNG 505

Db 345 GTYINCSRRGCKNDVCSFVDSCTCKAGMHGVDCSIRCPSTWGFGLTQCLNG 403

QY 506 ACNTLDGTCTCAPGMR 521

Db 404 LCHPVSSCQPGSGSR 419

RESULT 8

US-09-870-759-140

; Sequence 140, Application US/09870759

; Patent No. US20020177551A1

; GENERAL INFORMATION:

; APPLICANT: TERMAN, David S

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT OF NEOPLASTIC DISEASE

; FILE REFERENCE: 870759

; CURRENT APPLICATION NUMBER: US/09/870,759

; CURRENT FILING DATE: 2002-01-14

; PRIOR APPLICATION NUMBER: US 60/208,128

; PRIOR FILING DATE: 2000-05-30

; NUMBER OF SEQ ID NOS: 166

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 140

; LENGTH: 830

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-870-759-140

Query Match 23.4%; Score 843.5; DB 9; Length 830;

Best Local Similarity 36.2%; Pred. No. 7.4e-36;

Matches 158; Conservative 51; Mismatches 164; Indels 63; Gaps 15;


```

Db 302 ERTFARADRYCOCF---RGRCHPYDGTACACPEPTGRKTYCRKCPACAGTGLGCKRRRCGCG 358
OY 416 QNGADCDSDYTKC-TCAPBFGKIDCSTPCPIATYGINCSSRC-GCKNDYAVCSPYDGSCT- 472
Db 359 KQODPCTVAEGHCLTCEBGMNKTCKDQPCATGTYEGEGCSHRCPPORDGHACHVTKGKTR 418
OY 473 CKAGHGVDCSIRCPSTGTMWFGCNLTCCOLNGACNTIDGTCTCAPRGRGECCELPDODG 532
Db 419 CNAIGTIDRCETKCSNGTYGEDCAFVCADCGSGHCDPQSGRCLCSPGVHGPHCVNTCPBG 478
OY 533 TYGLNCAERCDCSHADGCHPTTGHG 557
Db 479 LHGADCAQACSC-HEDTCDPYTGAC 502

```

RESULT 12

```

US-09-866-050A-458
; Sequence 458, Application US/09866050A
; Publication No. US20030040471A1
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Ornust, Rene
; APPLICANT: Murison, James G.
; APPLICANT: Kumble, Krishnand D.
; TITLE OF INVENTION: Compositions for Isolated From Skin Cells
; FILE REFERENCE: 11000.1011c4U
; CURRENT APPLICATION NUMBER: US/09/866,050A
; CURRENT FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 725
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 458
; LENGTH: 296
; TYPE: PRT
; ORGANISM: Mouse
US-09-866-050A-458

```

```

Query Match 21.6%; Score 779; DB 9; Length 296;
Best Local Similarity 39.9%; Pred. No. 6,6e-33;
Matches 127; Conservative 27; Mismatches 118; Indels 46; Gaps 1;

```

```

OY 165 GACHCAGFRMGRCEDRCOEGTYNDCHQRCQCONGATCDHVTGECRCPPGTGAFCEDL 224
Db 1 GACYCPAGFLGADCSLACPGGRFPSCAHVTCGGAACDPVSTCTCPCPKTGHCBERG 60
OY 225 CPBGKHPQCEQRCPQNGGVCHVTGECSCPSGMGTVCQPCPEGRFGKNCQSECOCH 284
Db 61 CPDRFPGKCEHCKACRNGGLCHATNGSCSCLPMWGPCHCHACPARGYGAACLLCESQ 120
OY 285 NGCTCDATGQCHSPYTGRCODECPVGTGYVLAETCCVNGKCYHVSAGALCEAG 344
Db 121 NNGSCEPTSGALCGPFGYQACEDTFCPAGFHGSGCQRVCECQGAACDPVSGRCLPAG 180
OY 345 FAGERCEARLCPBGLYGIKCDKRCPLHNTSHCHPMGSEBACRPGMSGLYCNTECSPGF 404
Db 181 FRGO-----FCERGCKRPF 194
OY 405 YGACQOICSCONGADSDYTKCTCAPGFGKIDCSTPCPLGTYYGINCSSRCGCKNDYAVC 464
Db 195 FGDGCLQGNCPFTGVPDPIISGLCLCPRGRAGTTCDLDCRGRGRPGALRCDCGGGADC 254
OY 465 SPVDGSCCTCAGHGVDC 482
Db 255 DPISGQCHCVDSYTGPTC 272

```

```

RESULT 13
US-09-866-050A-192
; Sequence 192, Application US/09866050A
; Publication No. US20030040471A1

```

```

; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Ornust, Rene
; APPLICANT: Murison, James G.
; APPLICANT: Kumble, Krishnand D.
; TITLE OF INVENTION: Compositions for Isolated From Skin Cells
; FILE REFERENCE: 11000.1011c4U
; CURRENT APPLICATION NUMBER: US/09/866,050A
; CURRENT FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 725
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 192
; LENGTH: 299
; TYPE: PRT
; ORGANISM: Mouse
US-09-866-050A-192

```

```

Query Match 21.6%; Score 779; DB 9; Length 299;
Best Local Similarity 39.9%; Pred. No. 6,7e-33;
Matches 127; Conservative 27; Mismatches 118; Indels 46; Gaps 1;

```

```

OY 165 GACHCAGFRMGRCEDRCOEGTYNDCHQRCQCONGATCDHVTGECRCPPGTGAFCEDL 224
Db 4 GACYCPAGFLGADCSLACPGGRFPSCAHVTCGGAACDPVSTCTCPCPKTGHCBERG 63
OY 225 CPBGKHPQCEQRCPQNGGVCHVTGECSCPSGMGTVCQPCPEGRFGKNCQSECOCH 284
Db 64 CPDRFPGKCEHCKACRNGGLCHATNGSCSCLPMWGPCHCHACPARGYGAACLLCESQ 123
OY 285 NGCTCDATGQCHSPYTGRCODECPVGTGYVLAETCCVNGKCYHVSAGALCEAG 344
Db 124 NNGSCEPTSGALCGPFGYQACEDTFCPAGFHGSGCQRVCECQGAACDPVSGRCLPAG 183
OY 345 FAGERCEARLCPBGLYGIKCDKRCPLHNTSHCHPMGSEBACRPGMSGLYCNTECSPGF 404
Db 184 FRGO-----FCERGCKRPF 197
OY 405 YGACQOICSCONGADSDYTKCTCAPGFGKIDCSTPCPLGTYYGINCSSRCGCKNDYAVC 464
Db 198 FGDGCLQGNCPFTGVPDPIISGLCLCPRGRAGTTCDLDCRGRGRPGALRCDCGGGADC 257
OY 465 SPVDGSCCTCAGHGVDC 482
Db 258 DPISGQCHCVDSYTGPTC 275

```

```

RESULT 14
US-09-866-050A-332
; Sequence 332, Application US/09866050A
; Publication No. US20030040471A1
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Ornust, Rene
; APPLICANT: Murison, James G.
; APPLICANT: Kumble, Krishnand D.
; TITLE OF INVENTION: Compositions for Isolated From Skin Cells
; FILE REFERENCE: 11000.1011c4U
; CURRENT APPLICATION NUMBER: US/09/866,050A
; CURRENT FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 725
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 332
; LENGTH: 299
; TYPE: PRT
; ORGANISM: Mouse
US-09-866-050A-332

```

Query Match 21.6%; Score 779; DB 9; Length 299;
Best Local Similarity 39.9%; Pred. No. 6,7e-33;
Matches 127; Conservative 27; Mismatches 118; Indels 46; Gaps 1;

QY 405 YGEACQOICSCQNGADCDSTYTKCTCAPGFKGIDCSTPCPL 445
DB 238 FGPSCTLHCDGCGGADCDPVSGQCHVDYMGPTCRREGPL 278

Search completed: May 9, 2003, 12:02:22
Job time : 68.1495 secs

QY 165 GACHCAAGFRGRCEDRCBEGTYNDCHQRCQNGATCDHYTGECRCPPGYTGAFCEDL 224
DB 4 GACYCPAGFLCADCSLACPGREFGSPCAHVCTCGGAACDPVSGTCLCPKGTGHCERG 63
QY 225 CFPKHPGRCBQRCPCNGGCVCHVTGECSPSGMMGTVCQPCEPGRFGNCQEQCH 284
DB 64 CPQDRFGKCHRCACRNNGSCSCLGMMGPCEHACPAGRYGAACLLCECSQ 123
QY 285 NGCTDAATGQCHCSPGYTGRCDECPVGYGVLCAETCCQVNGKCYHVSACLCCEAG 344
DB 124 NNSCEPTSGACLCGPFYTGACEDTFCPAGHSGCCQRCVCEQOQAPCDPVSGRLCPAG 183
QY 345 FAGERCEARLCPBGLYIKCDKRCPCHELNTSHCSHMSGECACKPGWSGLYCNETCSPGF 404
DB 184 FRGQ-----FCERCKKRF 197
QY 405 YGEACQOICSCQNGADCDSTYTKCTCAPGFKGIDCSTPCPLGTGICNCSRGGKNDAYC 464
DB 198 FGDCGLQCCNCPVPCDPISGLCLCPPRAGTTCDLDCRRGRFPGCALRCDCGGGADC 257
QY 465 SPYDSCCTCKAGMHGVC 482
DB 258 DPISGQCHVDSTYTCPTC 275

RESULT 15
US-10-084-994-12
; Sequence 12, Application US/10084994
; Publication No. US20030023070A1
; GENERAL INFORMATION:
; APPLICANT: NI et al.
; TITLE OF INVENTION: Attractin-like polynucleotides, polypeptides, and antibodies
; FILE REFERENCE: PTO11P1C1
; CURRENT APPLICATION NUMBER: US/10/084,994
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/790,621
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: PCT/US00/23663
; PRIOR FILING DATE: 2000-08-29
; PRIOR APPLICATION NUMBER: US 60/151,348
; PRIOR FILING DATE: 1999-08-30
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 12
; LENGTH: 310
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-084-994-12

Query Match 20.0%; Score 721.5; DB 9; Length 310;
Best Local Similarity 44.5%; Pred. No. 5.5e-30;
Matches 125; Conservative 24; Mismatches 129; Indels 3; Gaps 2;

QY 165 GACHCAAGFRGRCEDRCBEGTYNDCHQRCQNGATCDHYTGECRCPPGYTGAFCEDL 224
DB 1 GPAAPLSSGRTATSPVPOGRFGRNCTHYGCGGACADPVITGTCCLPPGRACVHCERG 60
QY 225 CFPKHPGRCBQRCPCNGGCVCHVTGECSPSGMMGTVCQPCEPGRFGKNCQEQCH 284
DB 61 CPOHNFVGECHTSCRNNGSLCHANSNGSCGGLWTRHCELCAPPRITGAACHLECSCH 120
QY 285 NGCTDAATGQCHCSPGYTGRCDECPVGYGVLCAETCCQVNGKCYHVSACLCCEAG 344
DB 121 NNSCEPATGTCRCGPGFYGQACEHPCPGPHGAGCGQLCRQCHGAPCDPVISGRCLPAG 180
QY 345 FAGERCEARLCPBGLYIKCDKRCPCHELNTSHCSHMSGECACKPGWSGLYCNETCSPGF 404
DB 181 FHGHCE-RGCEPSFEGCHQRCDC--DGAPCDPVITGLCLCPPGRSGATCNLDCCRQ 237

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4 p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 9, 2003, 11:47:55 ; Search time 16.9757 seconds

(without alignments)
3318.560 Million cell updates/sec

US-10-092-390-4

Sequence: 1 MYSLNSCLSFICLLICHLMI.....HDSVCAEGRMGNCLPCY 586

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database 8
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1372.5	38.1	1574	2 T13954	MEGF6 protein - ra
2	1322	36.7	1620	2 T27283	hypothetical prote
3	1284.5	35.7	1111	2 T26872	hypothetical prote
4	719	20.0	2524	2 A35844	notch protein - Af
5	717	19.9	2437	2 S42612	transmembrane prot
6	697	19.4	2321	2 S78549	notch3 protein - h
7	693	19.2	2531	2 A46019	Notch-1 protein -
8	687	19.1	1203	2 A49175	Notch B protein -
9	685.5	19.0	2318	2 S43506	notch 3 protein -
10	685.5	19.0	2471	2 A49128	cell-fate determin
11	682	18.9	3566	1 A40701	tenascin-X precurs
12	681.5	18.9	2531	2 T31070	notch homolog - se
13	677	18.8	2703	1 A24420	notch protein - fr
14	676.5	18.8	4006	2 T09070	probable tenascin
15	675	18.7	2531	2 S18188	notch protein homo
16	672.5	18.7	4135	2 T42629	tenascin-X - Bov
17	664.5	18.5	1064	2 A40136	fibropellin Ia - s
18	659.5	18.3	2555	2 A40043	notch protein homo
19	658	18.3	1964	2 T09059	notch4 - mouse
20	648.5	18.0	2352	2 T30201	Notch homolog prot
21	644.5	17.9	2201	2 A32160	tenascin-C - human
22	640.5	17.8	2019	1 J01322	tenascin precursor
23	636	17.7	2130	1 A32230	tenascin precursor
24	631.5	17.5	2139	2 A35672	crumbs protein - f
25	616	17.1	1220	2 A56136	tenascin precursor
26	611	17.0	1746	1 S19694	hypothetical prote
27	593.5	16.5	3672	2 T23433	probable laminin a
28	593.5	16.5	3704	2 T37316	gene serrate prote
29	587	16.3	1408	2 S16148	

30	586	16.3	1722	2 E89753	protein E11C7.4 (1
31	586	16.3	3635	2 T10053	laminin alpha 5 ch
32	580.5	16.1	861	2 A48825	Notch homolog Mote
33	577.5	16.0	1801	1 MMRIS	laminin beta-2 cha
34	576.5	16.0	3712	2 S18253	laminin alpha-1 ch
35	565.5	15.7	2823	2 T23064	hypothetical prote
36	565.5	15.7	2823	2 F87908	protein T22A3.8 (1
37	565.5	15.7	3102	2 T43291	laminin alpha chal
38	562.5	15.6	647	2 A43802	tenascin - eastern
39	561	15.6	1798	2 S53869	laminin beta-2 cha
40	561	15.6	3106	2 S53868	laminin alpha-2 ch
41	560.5	15.6	1429	2 S06434	homeotic protein 1
42	556	15.4	833	2 S19087	gene Delta protein
43	552	15.3	832	2 A31246	neurogenic protein
44	552	15.3	880	2 S00670	neurogenic repeat
45	550	15.3	1797	2 A53677	laminin beta-2 cha

ALIGNMENTS

RESULT 1

T13954
MEGF6 protein - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 21-Jul-2000
C/Accession: T13954
R:Nakayama, M.; Nakajima, D.; Nagase, T.; Nomura, N.; Seki, N.; Ohara, O.
Genomics 51, 27-34, 1998
A:Title: Identification of high-molecular-weight proteins with multiple EGF-like motif
A:Reference number: Z14126, M0ID:98360089, PMID:9693030
A:Accession: T13954
A>Status: preliminary; translated from GB/EMBL/DBD
A:Molecule type: mRNA
A:Residues: 1-1574 <NA>
A:Cross-references: EMBL:AB011532; NID:93449293; PIDN:BA32462.1; PID:93449294
A:Experimental source: strain Sprague-Dawley; brain
C:Genetics:
A:Gene: MEGF6

Query Match	38.1%	Score 1372.5;	DB 2;	Length 1574;
Best Local Similarity	41.3%	Pred. No. 7.4e-69;		
Matches 229;	Conservative 56;	Mismatches 198;	Indels 71;	Gaps 9;
QY	94	CCPEFYSGEMCVPHCADKVHG-----RCIAAPV-----TCOEPGNGTNC	136	
DB	808	CLPEFYGS-----RCDDTSAGWYGTGICRCACANDGHCDPTTGRSCAPGTGLSCQ	861	
QY	137	SACDGDHMGPHCTSRCCCKNG-ALCNPTGACHCAAGFRGRCEDRCOGTYGNDCHQC	195	
DB	862	RACDSGHMGPDCTHPCNCSAGHNCDAVSLGCLCEAGYEGPRGQSCRGTYGSCDKC	921	
QY	196	QCONGATCDHYTGRCRCPPYTGAFCD-----LC 225		
DB	922	RCEGACDHYSGACTCPAGWRGSCFCHACPAGEFGIDCDACNCSAGACDAVTCSCIC	981	
QY	226	PRGHGPGQCEQRCP-----CONGVCHHTGTGRCSCPSPGMYGVCQPCPEGR	272	
DB	982	PAGRWGRCAOSCPPLFTGLNCSQICTCFMGASDSTYTGHCAPGMWGPCTCDACPPGL	1041	
QY	273	FGKNCSECCCHNGTCDATATGQCCHSCSPGYTGRCDECEVGTGVCAETCCQVNGKC	332	
DB	1042	YGRNQCCHSCCLRNCRGCRDPLIGCTCEBWTGLACENBCLPRHNAACQALNCSLHGIC	1101	
QY	333	YHVSAGCLCEAGFAGECEARLCEGLYGIKCDRCPCHELEHTSCHPMSEACAPRGS	392	
DB	1102	DLRTGHCLCPAGWTGDKQSS-CVSGTFGVNCEHCAC--RKASCHHTVAGACFCPPGWR	1158	
QY	393	GLVCNCTSPGFGEAGCQICSCQNGADCSYNGKCTCAGCFGIDCSTPCPGTYGIC	452	
DB	1159	GPHCEACPRGWFEACQACRLCPTNASCHHTVGECCPCPGFTGLSCQACQCPETFEKDC	1218	
QY	453	SSRGCCKNDA-VCSPYDGSCTCKAGWGHGVDCSIRCPSTGTFGFCNLCCQLNGAGCWTLD	511	


```

Oy 182 CEQGYGNDCHQRCQCCGATC-DHYTG-ECRCPEGTGACEDLCPPKGGPOCEORCP 239
Db 790 I-----NECASN-PCLNQSGCIDVAVGFKCNMLPTGYEVEENTLAP-----CSPR-P 835
Oy 240 CONGVCHH-----VTGECSPSGMGTVCQGPCEPGRFGKNCOSBC-----OCHNGSTQAA 292
Db 836 CKNGVCHESDPSFSCNPGAGMOCCEVDI-----NECVNPNCTNGVCENL 885
Oy 293 TG--QCHCSPGYTGEBRCQ---DECPVGTGYVLAETCCVNGKCY-HVSG-ACLCENAGF 345
Db 886 RGGFCQRCNPNFTGALCENDIDC-----EPNCSNGVQCDRVNNGFVCLAGF 935
Oy 346 AGEKCEARL-----CPEGYXGKCDKRCRCHLEHNSCHP- 380
Db 936 RGEKCAEDIDCVSAPCNGNCTDCVNSYTCSCPAGESSGINCENIPDCTES--SCFNG 993
Oy 381 -----MSGECAKPGMSGLYC---NE-----TCSPGFYGEA 408
Db 994 GTCVDGJSSFSQVCLPFTGTGNYCQHDVNECDSPQNGSCQDGTGTYKCTCPHGTGLN 1053
Oy 409 CQOI---CS---CONGADC--DSYTGKCTCAPFRKIDCSTP-----442
Db 1054 CQSLVRMCDSSPCXKNGSCWQOGASFCCOCASGWTGTYCDVPVSCEVAARQGVSAVL 1113
Oy 443 -----CPLGTGINSRSG-----CKNDAYCSPVDS--SCTCRAGW 477
Db 1114 CRHAGCCVDAGNTHLCRCQAGTSTYCOEVDDECQPNPCAGATCTDYLGTSCCEVPGY 1173
Oy 478 HGVDCS-----IRCPSGTWFGCNL---TC-----499
Db 1174 HGMCSEINECLSQPCQNGSTCIDLVNTYKSCPRGTQGVHCEIDIDCSPSYDPLTGE 1233
Oy 500 -OCLNGACNTLDG--TCTCAPWRGKECE-----LRQ-DGTGLNCAE-----RC 542
Db 1234 PRGNRGRCVDRVGGYCCVCPAGFVGRCEQVNECLSDPDPSGSY--NCVQLINDFR 1291
Oy 543 DCSHA-----DCCHPT-----TGH-----CRCLPGMSGVCD-----569
Db 1292 ECRFGYTGKRCETVFNCKXDPCKNKGTCANVASTKHGICYCKCPGTSGSSCEYDSQSG 1351
Oy 570 -----SVCAEGRWGPC 581
Db 1352 SLRCRNGATCVSHLSPRC 1370

```

RESULT 6

578549 notch3 protein - human

C:Species: Homo sapiens (man)

C:Date: 24-Jul-1998 #sequence_rev1sion 24-Jul-1998 #text_change 02-Aug-2002

C:Accession: S78549; S71825

R:Jourel, A.; Tournier-Lasserre, E.

submitted to the EMBL Data Library, April 1997

A:Reference number: S78549

A:Accession: S78549

A:Molecule type: mRNA

A:Residues: 1-2321 <JOU1>

A:Cross-references: EMBL:U97669; NID:q2668591; PIDN:AB91371.1; PID:q2668592

R:Jourel, A.; Corpechot, C.; Ducros, A.; Vahedi, K.; Chadrillac, H.; Mouton, P.; Alamiowitc

x, M.M.; Weissenbach, J.; Bach, J.F.; Bousset, M.G.; Tournier-Lasserre, E.

Nature 383, 707-710, 1996

A:Title: Notch3 mutations in CADASIL, a hereditary adult-onset condition causing stroke

A:Reference number: S71825; MUID:97032728; PMID:8878478

A:Accession: S71825

A:Status: nucleic acid sequence not shown

A:Molecule type: DNA

A:Residues: 67-113;138-194;268-333;G',335-346;536-613;716-765;1240-1279;1815-1888 <JOU2

A:Cross-references: EMBL:U97669

C:Genetics:

A:Gene: notch3

A:Map position: 19p13.1

C:Function:

A:Description: may be involved in pathogenesis of CADASIL, causing a type of stroke and

```

C:Superfamily: notch protein; ankyrin repeat homology; EGF homology; laminin-type EGF.
C:Keywords: tandem repeat; transmembrane protein
F:123-155/Domain: EGF homology <EGX1>
F:162-194/Domain: EGF homology <EGX1>
F:240-271/Domain: EGF homology <EGX2>
F:318-349/Domain: EGF homology <EGX>
F:473-504/Domain: EGF homology <EGX3>
F:853-884/Domain: EGF homology <EGX3>
F:928-959/Domain: EGF homology <EGX4>
F:1070-1126/Domain: laminin-type EGF-like homology <LEGS>
F:1838-1870/Domain: ankyrin repeat homology <AN1>
F:1871-1903/Domain: ankyrin repeat homology <AN2>
F:1905-1937/Domain: ankyrin repeat homology <AN3>
F:1938-1970/Domain: ankyrin repeat homology <AN4>
F:1971-2003/Domain: ankyrin repeat homology <AN5>

Query Match 19.4% Score 697; DB 2; Length 2321;
Best Local Similarity 25.28; Pred. No. 1.8e-31;
Matches 226; Conservative 61; Mismatches 249; Indels 360; Gaps 51;

Oy 5 LNSCLIS-----FICLLCHWIGTASPLNLED-----PNVC-SHMEY 40
Db 432 VNECLSGPCRNQATCLDRIGQFTICMAGFTGYCEVDIDECQSSPCVNGVCKDRNGF 491
Oy 41 SVTYQESYPPHFDQIYITSC--TDILMFKCTRHRVYFRAYRHGERTMRKSCCPGF 98
Db 492 SCTCPSEFSSTCQLDVDECASTPCRNARAKVDQPDQY-----ECRCAEGF 537
Oy 99 YESGEMC--VPHCA-DKCVHRC---IAPTCQCEGMGTNCSACDDDHMPHCTSR 151
Db 538 -EETLCDRVNDQSPRCHHRCVVDIATSCACAGYGTRESQYD-----ERSQ 589
Oy 152 CQCKNGALCNPTTG--ACHCAEGRWRC--DRCEQ--TYG--NDCHQR--CQCONG 200
Db 590 -PCRHHGKCIDLVKYLRCRPSGTYGVNIDCASNPCTFCVCHDGINRYCVCQPG 648
Oy 201 AT---CDHYTGECRCPPGYGARGED-----LCPRKGRPQC---EQRC--PCQNG 244
Db 649 FTGPLCAVEINECASSPCGEGSCVDENGFRCLCPGPSLPLCLPSHPCAHPCSH-G 707
Oy 245 VCHHVTG--ECSCPSGMGTVCQO-----PCEGRFGK 275
Db 708 ICYDAPGFRVCYCEPBGSGRCSSILARDACESQPCAGTCSDDGFGFCTCPGYQGR 767
Oy 276 NCS--QEC---QCHNGSTCAATGQ---CHCSPGYTGRCQ---DEC---PVGTYVLC 320
Db 768 QCELLSPCTPNPCHEHGRCESAPQPLVYCSQPGWQGRQOVDECAQAPACGPHCI-C 826
Oy 321 AE-----TCQ-----CVNCGKCHVSG--ACLCRAGFRGCEA 352
Db 827 TNLAGFSCTCHGGYTGSPCDQDINDCPNCLNGGSCQDGVGSFSCSLPFGAFPRC-A 885
Oy 353 R-----SGEACRPGMSGLYCNF-----LCPEGLYGKICDKRCPCHLEHNSCHPM-----381
Db 886 RDVDECLSNPCGPTCTDHAVSFTCTCPPGYGGHCGDLP-----DSSPSSCFNGGT 938
Oy 382 -----SGEACRPGMSGLYCNF-----TCSPGFYGBAQ 410
Db 939 CVDGVNSFSLCRBRGTGAHCQHEADPCLSRPLHGVCSAARPRGCTLESFTGPOCQ 998
Oy 411 QI---CS---CONGADDSVTGKCTCAPFRKIDC---SNP-----442
Db 999 TLVDKMSRQPCQNGRCVQVGAICLCPGWSGRICLIRSLPCBAAQAIVRLDQLCQAG 1058
Oy 443 -----CPLGTGINSRSG-----CKNDAYCSPVDS--CTCRAGHGV 481
Db 1059 GQCVDEDSHYCYCPGPRGSHCEQEVDPCLAPQCHGRCRQYMGYMECLPYNGDN 1118
Oy 482 CS-----IRCPSTWFGCNLT---C-----OCLN 503
Db 1119 CEDDVDECAQPCQHGSCIDLVARYLSCSPGTLVGLCEINEDDCGPGPLDSCGRCLH 1178
Oy 504 GACNNTLDG--TCTCAPWRGKECEL---PCQDGTGLNCAEKGDCSHDGHCHPTTG---555

```

```

Db 1179 NGCTVDLGGFRCCTPCPYGTGIRCEADINECRSGA-----CHAATRTCLADPPGGGF 1230
OY 556 HCRCLPGMSGVHCDY-----CAMGRGPGNC 581
Db 1231 RCLCHAGFSRRCQTVLSPCESQPCQHGCGRSPGPGGLTTCCHCAQPTWGPGR 1286

RESULT 7
A46019
Notch-1 protein - mouse
N:Alternate names: notch protein
C:Species: Mus musculus (house mouse)
C:Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 20-Sep-1999
C:Accession: A46019; S25144
R:del Amo, F.; Gendron-Maguire, M.; Swiatek, P.J.; Jenkins, N.A.; Copeland, N.G.; Grid
Genomics 15, 259-264, 1993
A:Title: Cloning, analysis, and chromosomal localization of Notch-1, a mouse homolog of
A:Reference number: A46019; MUID:93194170; PMID:8449489
A:Accession: A46019
A:Status: not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-2531 <DEL>
A:Cross-references: GB:211886; GB:S47228; NID:q288502; PIDN:CAA77941.1; PID:q288503
A:Note: sequence extracted from NCBI backbone (NCBI:127318)
R:Franco del Amo, F.; Smith, D.E.; Swiatek, P.J.; Gendron-Maguire, M.; Greenspan, R.J.;
submitted to the EMBL Data Library, April 1992
A:Description: Expression pattern of Notch, a mouse homolog of Drosophila Notch, suggest
A:Reference number: S25144
A:Accession: S25144
A:Molecule type: mRNA
A:Residues: 1551-2108, 'Q', 2110-2114, 'ALP', 2118-2170 <FRA>
A:Cross-references: EMBL:Z11886
C:Genetics:
A:Gene: notch-1
A:Map position: 2
A:Note: proximal region of chromosome 2
C:Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homology
F:106-18/Domain: EGF homology <EGF1>
F:144-175/Domain: EGF homology <EGF1>
F:222-254/Domain: EGF homology <EGF2>
F:261-282/Domain: EGF homology <EGF2>
F:339-370/Domain: EGF homology <EGF3>
F:416-449/Domain: EGF homology <EGF3>
F:456-487/Domain: EGF homology <EGF4>
F:494-525/Domain: EGF homology <EGF5>
F:533-563/Domain: EGF homology <EGF6>
F:607-638/Domain: EGF homology <EGF7>
F:682-713/Domain: EGF homology <EGF8>
F:757-788/Domain: EGF homology <EGF9>
F:795-826/Domain: EGF homology <EGF10>
F:873-904/Domain: EGF homology <EGF11>
F:911-942/Domain: EGF homology <EGF12>
F:949-980/Domain: EGF homology <EGF13>
F:987-1018/Domain: EGF homology <EGF14>
F:1025-1056/Domain: EGF homology <EGF15>
F:1063-1094/Domain: EGF homology <EGF16>
F:1149-1180/Domain: EGF homology <EGF17>
F:1187-1218/Domain: EGF homology <EGF18>
F:1233-1264/Domain: EGF homology <EGF18>
F:1352-1383/Domain: EGF homology <EGF19>
F:1391-1425/Domain: EGF homology <EGF19>
F:1497-1528/Domain: EGF homology <EGF19>
F:1517-1548/Domain: ankyrin repeat homology <AN1>
F:1549-1581/Domain: ankyrin repeat homology <AN2>
F:1583-2015/Domain: ankyrin repeat homology <AN3>
F:2016-2048/Domain: ankyrin repeat homology <AN4>
F:2049-2081/Domain: ankyrin repeat homology <AN5>

```

```

Db 624 SYLCLLKGTTGPNCENILDD---CA-----SNPCDS---GTCLDKIDGEC 664
OY 70 TRHRVYTAIRRHGKTYRRRSQCCPGFYEGENC---VPRCA----- 110
Db 665 A-----CEPGY---TGSNCNVNIDECASSPCHNGTCEDGIAG 699
OY 111 -----DKVHGRC---IAPNTCOCEPGMGGTC---SSACDG 141
Db 700 FTGCRPEGYHPTCLSEVNECNSNPCIHGACDGLNGKCDCAPEMSGINDINNEDS 759
OY 142 DHMGPHCTSRQCKNGKALCNPTTG---ACHCAFGWRGRCEDRCGTYGNDCHORCOQON 199
Db 760 N-----PCVNGGCKDKMTSGVYCTCREGFGSPNCO-----TINIECASN-PCLN 802
OY 200 GATC-DHYTG-ECRCPRYGTAFCDLCPGKHGRCQECORRCQNGVC---HHVTGEC 253
Db 803 GGTCTIDVAGYKCNCPDLYTGATCEVLAPE-----C-ATSPCKNSGVCKESEDYESFSC 855
OY 254 SCPSGMGWTC-----GQPCPEGR-----FGKNC---QRCQ--- 282
Db 856 VCPFGWQGTCEVDINECVKSPCRHAGASCQNTNGSYRLCLQAGTYGRNCESDIDCRNP 915
OY 283 CHNGSTCDAA--TGQCHSPGYTGERCODE-----CPVGT 315
Db 916 CHNGSGCTDGINFARCDLPGHGAFCEDINECASNPCQNGANCCTDQVSYTCGPYGF 975
OY 316 YGVLCALP-----CQCVNGKCIYVG---ACICEGFAFEREARLCPGLYGT-KC 364
Db 976 NGIHENNTPDCTESSCENGTC--VDGINSFTCLCPPTGSSYCO-----YDVNEC 1025
OY 365 DKRCCHLENTHSCHPMSG--ECACKPGMSGLYCNE-----TCSPGFYGEACQOI----- 412
Db 1026 DSR-PCLHGCT--CODSYGTIKCTCPQGTGILNCQNLVRCMDSAPCKNGRCQNTQYH 1082
OY 413 CSCQN---GADCDVYTGKCTCAPRGKIDCSPTCPGLGYGIN-----CSSRCG----- 457
Db 1083 CECRSGWGVNCDVLYSCEVAQKRGIDVTLQCHGGILCVDGDKHVCCHCQAGYTGYSVC 1142
OY 458 -----CKNNAVSPVDG--SCICKAGHWGVDS----- 483
Db 1143 EDEVDKSPNCPQNGATCTDYLGGFSCKCAVAGHSNCSSEINELSLQPCQNGTCIDLT 1202
OY 484 ----TRCSGTWGFSGCNTL--C-----OCLNGACNTLDG--TTCAPAGWGE 523
Db 1203 NSYKSCPRGTOGVHCEINVDCHPPLDPASRPFNNGTCVVDYGVGTCPCPGFVGE 1262
OY 524 KCE-----LPCOD-GTYGLNCAER-----SHADG----- 550
Db 1263 RCEGVNCLSNPCDPRGTQ--NCVQRVNDFHCRCRAGHTGRCSYVINGCRPKCKNG 1320
OY 551 -----HFTTGH-CKCLGMSGVHDS-----VCACGRMGPNP----- 581
Db 1321 VCAVASNTARGFTCRCPGFEGATCENDARTGSLRCLNGTCTIGSPSPCTCLGSEFTG 1380
OY 582 -----SLPCY 586
Db 1381 PECQPASSPCVGSNPCI 1398

RESULT 8
A49175
Notch B protein - mouse (fragment)
N:Alternate names: Notch homolog
C:Species: Mus musculus (house mouse)
C:Date: 21-Jan-1994 #sequence_revision 05-Jan-1996 #text_change 20-Sep-1999
C:Accession: A49175; PHL570; S32113
R:Lardelli, M.; Lendahl, U.
Exp. Cell Res. 204, 364-372, 1993
A:Title: Notch A and Notch B-two mouse Notch homologues coexpressed in a wide varie
A:Reference number: A49175; MUID:93178563; PMID:8440332
A:Accession: A49175
A:Status: preliminary; nucleic acid sequence not shown

```

OY 10 SFICLLCHWIGTASPLNLEDPNVCSHWESYSVTQESYPPHFDQIYTTCTDILNMFRC 69

Query Match 19-2%, Score 693; DB 2; Length 2531;
 Best Local Similarity 25-3%, Pred. No. 3-2e-31;
 Matches 217; Conservative 71; Mismatches 206; Indels 364; Gaps 51;

A: Molecule type: mRNA
 A: Residues: 1-1203 <LNR>
 A: Cross-references: EMBL: X68279; NID: g287989; PID: CAA48340.1; PID: g287990
 A: Experimental source: embryo
 A: Note: sequence extracted from NCBI backbone (NCBI: 126158)
 C: Comment: This protein has many EGF repeats and 11n-12/Notch repeats.
 C: Comment: This protein is one of the neurogenic proteins controlling the decision between C: Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homology
 F: 143-174/Domain: EGF homology <EGF1>
 F: 482-513/Domain: EGF homology <EGF1>
 F: 560-591/Domain: EGF homology <EGF>
 F: 674-705/Domain: EGF homology <EGF2>
 F: 712-743/Domain: EGF homology <EGF3>
 F: 836-867/Domain: EGF homology <EGF3>

Query Match 19.1%; Score 687; DB 2; Length 1203;
 Best Local Similarity 24.8%; Pred. No. 4,3e-31;
 Matches 221; Conservative 78; Mismatches 245; Indels 348; Gaps 56;

```

3 ISLNSCLSFICL---LICH-----WIGTASPLNE--DPNVCSSM-----ES 39
214 IDIDCSSTPCLNGAKCIDHPNGYECOCATGFTGILCDENIDNCDDPCHHGQCCDIDS 273
40 YSTVQESYPRHF--DOI--YYS-----CTDILNMEKCT-----RHYVSY---- 76
274 YTCINPMTWGAICSDQIDECYSSPCLDNGRCIDLNVGYQCNCQPGTSLCEINPDCA 333
77 RTAYRAG--EKTMYRRKSOCPCGF-----YESGEMCV----- 106
334 SNCHMGVYVDGINRYSVCSPFTGRCIDIDECASNCRKAGTICINDVNFRCICPE 393
107 ----PHC-----ADKCVHGR---IAPNTQCEPBGWGTNCSSACDGHMGPCTSR 151
394 GPHHPSCYSOVNCLNPNCTHGNCGLSGYKCLCDAGWGVNCE--VDKN-----ECLSN 447
152 COCKNALCNPIIGA--CHCAAGFRGRCE--DRC-----EGTYGNCH--QRCCO-- 198
448 -PCQNGCTNNLVNGYRCKTKGFKGICVNVNIDECASNCLNCGTCFDDVSGYTGCM 506
199 --NGATCDHYTGRCRPRGTGAFCE-----LCPRKHPGPOE--QRC--PCQ 241
507 PTYGRKNCQYTLARCSNPNCNAVKEARNFESFSLCAFGMGKKTVDVDCISKPCM 566
242 NGCVCHHVTGE--CSCPBGMTGVCQPCPEGRFKNCSOE--OCHNGTGC--DAATG 294
567 NNVCHNTGSSYVCECPRGFSGMDCEEDI-----NDCLANPCQNGSCVDHNTF 616
295 QCHSPGYGGERCODE-----CPVGTGYVLC--AETC-- 324
617 SCQCHPFTGDKQTDQTNNECLSEPCKNNGTCSYVNSYTCCTCPAGFHVAGCENNIDCTE 676
325 -QCVNGKCYHVSG---ACICEAGFAG-----ERCEAR-----LC 355
677 SSFNGTGC--VDGINSFSGLCVGTGPRCLHDINECSSNPCLNAGTCTVDGLTYRCIC 734
356 PEGLYVICKD-----KRCPCHELNTHSCHPMSEGCACRFGWSGLYCNE----- 398
735 PLGYTGKNCQTLVNLCSRSCKKNGTCVOKARPHCLCPRGMDGAYCDVLVNSCKAALQ 794
399 -----TSPGFTYGACQO-----ICS-----CQNGADDSTVG-- 426
795 KGVVPEHLCOHSGICINAGNTNHCQCPLETGYSYCEQULDEECASNPCHOMATNDNFIQY 854
427 KCTCAGFGKIDSTPCPLGTGYGINSRQCNDAVCSYVDSCTCKAGMHWVDSTIR-- 486
855 RECCEVGYGVNCE-----YEVDCQNPONGGTCTIDLNVHFKCS-----C 896
487 PSGTWFGG--NL--TC-----QCLNGAGC--NTLDG--TCTCAPGWRGECSE-----LPC 529
897 PPTRIELCEENIDECAGGPHCLNGGQVDRIGYTCRCLPFAAGRECEBDINECLSNPC 956
530 -QDGTGLNCAE-----RCDCSHA-----DGC-----HPTTGHC 557

```

DB 957 S5EGS--LDCVOLKNYNYNCICRSAPFTGRHCEFLDYCPQKPCNLNGTCAVANMMDGFTIC 1014
 QY 558 RCLPGMSGVCHDSVCAEGRW-----GNPC-----SLPC 585
 DB 1015 RCPGFSGARCCSSCGGVKCRGECIHTDSGFRCLNPKDCESGCASNPC 1066

RESULT 9

S45306
 notch 3 protein - mouse
 C: Species: Mus musculus (house mouse)
 C: Date: 20-Feb-1995 #sequence, revision 20-Feb-1995 #text, change 02-Aug-2002
 C: Accession: S45306
 R: Lardelli, M.; Dahlstrand, J.; Lendahl, U.
 Mech. Dev. 46, 123-136, 1994

A: Title: The novel Notch homologue mouse Notch 3 lacks specific epidermal growth fact
 A: Reference number: S45306; MID: 95001556; PMID: 7918097
 A: Accession: S45306

A: Status: preliminary
 A: Molecule type: mRNA
 A: Residues: 1-2318 <LNR>
 C: Superfamily: notch protein; ankyrin repeat homology; EGF homology
 F: 163-195/Domain: EGF homology <EGF1>
 F: 474-505/Domain: EGF homology <EGF1>
 F: 854-885/Domain: EGF homology <EGF2>
 F: 1839-1871/Domain: ankyrin repeat homology <AN1>
 F: 1872-1904/Domain: ankyrin repeat homology <AN2>
 F: 1906-1938/Domain: ankyrin repeat homology <AN3>
 F: 1939-1971/Domain: ankyrin repeat homology <AN4>
 F: 1972-2004/Domain: ankyrin repeat homology <AN5>

Query Match 19.0%; Score 685.5; DB 2; Length 2318;
 Best Local Similarity 24.4%; Pred. No. 7.9e-31;
 Matches 216; Conservative 59; Mismatches 195; Indels 417; Gaps 48;

```

94 CCEGTESEGEVYPHCADKCVHGRCTAPNT-----CQCEPBGWGTNCSSACDGHW 144
225 CLPGE--EGQNCSEVN--VDPCRGHRCCLNGTGVNVYTNQCQCPPEWTGQCTEDVD-- 277
145 GPHCTSRQO-----CKNGALCNPIIG--ACHCAAG-----FRGWRCE 179
278 -----ECQLDPNACHNGGTCFNLGHSVCVNGMTGESCSQNIIDCATAVCFHGATCH 331
180 DR-----CEQGTYNDCN--ORC--OQNGATCD--HYTGE--CRCPGYTGAFCE 223
332 DRVASFYCACPMKGTGLCHLDACVSNPCHEDIAIDTNVSRALCTCPRGFTGACDQ 391
224 -----LCPRGHNGPOE----- 235
392 DVDECSIGANPCHEHLGRCVNTQGSFLQCCGCGYTGPRCEVDNECLSGPCRNQATCIDRI 451
236 -----ORCPONGGYC--HHVYG--ECSCPGMMGTVC----- 264
452 GQFTCIAGFTGYTEVDIDECOSSPCVANGVCKDRVNFESCTCPBSGFSMCLDVDE 511
265 -----GQP-----CPEGRFKNCSQ--EQ--CHNGGTGDA--ATGQCIC 298
512 CASTPCRNKAKCYDQPDGYEBCRCABGFEGTLCERANNDDCSPDCHHGRVGLASFSAC 571
299 SPGYTGERCODE-----CPVGTGYVLC-----AETCQ 325
572 APGYTIRCSGYVDECRSQPCRYGKCLDLVDLYLRCRPGTGVNCEVNIIDCASNPCY 631
326 --CVNGKCYHVSAGCLCEAGFAGERCARL-----PREGLYGKCDKRCR 369
632 FGVCRDGINRYD---CYQCPGFTGPLCNYEINIECASSPGEGSGCVDEGNGFHC--LCP 665
370 -----CHLEHTS--CHPMSG--ECACKPGMSGLYCNE----- 398
686 PGLSLPLCLPANHPCAKHRCPSHGVCHDABRGFCVCEPFGHSGRCSQSLAPDACESQPCQ 745
399 -----TSPGFTYGACQOI--CS--CQNGADDSTVGK--CTCAGGFKG 436

```

```

Db 746 AGTCTSDGIGFRCCTCAGGFGHCEVLAESPCTPSLCEHGHGCESDPRLATVCSCPFGMG 805
Qy 437 -----IDCTPCPLGTGTCSS-----RCGCKNDAY-----CSP----- 466
Db 806 PRCDVDDECAAGSPGCGPHG--TCNTLPGNFRCICRGYTGAFPCDDIDDCPNCLHGS 864
Qy 467 -VDG-----SCCKAGMHWVDC-----SIRCSGTGFGCNL 497
Db 865 CQDVGSFSGCLDGFAGPRCARDVDECLSPGPGTCTDTHVAFSTCACPGYGGFCEI 924
Qy 498 -----TCQCLNGACNTLIDG-----TCTCARGMGEKC----- 525
Db 925 DLPCSPSSSCFNNGTC--VDGVSFSCLCRPGYGTHTQYEAADPCFSRCLHAGICNPTH 982
Qy 526 ---ELPCODGTGTLNCAERCD-----CSHADGHPPTGHCRCPLGMSGVHCD----- 569
Db 983 PGFECTRGFTGSCQCNFVDMCSQAPCQNGRCVOTGAYCICPBGMSGRCLDIQSLPCT 1042
Qy 570 -----SYCABGRMGPNCSL---PC 585
Db 1043 EAAAMGVRLLEQLCEGGKCLDKGRSHYCVCPBGRTGSHCEHYDPC 1089

RESULT 10
A49128
cell-fate determining gene Notch2 protein - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 21-Jan-1994 #sequence_revision 18-Nov-1994 #text_change 02-Aug-2002
C:Accession: A49128
R:Meinmaster, G.; Roberts, V.J.; Lemke, G.
Development 116, 931-941, 1992
A:Title: Notch2: a second mammalian Notch gene.
A:Reference number: A49128; MUID:93202015; PMID:1295745
A:Accession: A49128
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-2471 <WEI>
A:Experimental source: Schwann cell
A>Note: sequence extracted from NCBI backbone (NCBI:P:127811)
C:Superfamily: notch protein; ankyrin repeat homology; EGF homology
F:264-295/Domain: EGF homology <EGX1>
F:799-830/Domain: EGF homology <EGF1>
F:877-908/Domain: EGF homology <EGX2>
F:1029-1060/Domain: EGF homology <EGF>
F:1067-1098/Domain: EGF homology <EGX3>
F:1153-1184/Domain: EGF homology <EGX3>
F:1191-1222/Domain: EGF homology <EGX4>
F:1876-1908/Domain: ankyrin repeat homology <AN1>
F:1909-1941/Domain: ankyrin repeat homology <AN2>
F:1943-1975/Domain: ankyrin repeat homology <AN3>
F:1976-2008/Domain: ankyrin repeat homology <AN4>
F:2009-2041/Domain: ankyrin repeat homology <AN5>

Query Match 19.0%; Score 685.5; DB 2; Length 2471;
Best Local Similarity 24.8%; Pred. No. 8,2e-31;
Matches 218; Conservative 83; Mismatches 240; Indels 339; Gaps 55;

Qy 3 ISLNSCSFICL---LICR-----WIGTASPLNE--DPNVCSHM-----ES 39
Db 531 IDIDDCSTPCLNCAKCIDHPNGYECOCATFTGLCENIDNCDPDCGHCQCDGDS 590
Qy 40 YSVAVQESYRHPF--DQI--YYS-----CTDILNFKET-----RHRYSY---- 76
Db 591 YTCICNPGYMGALCSDDQDEICSSPCLNDGRCIDLVNMYQCNCPGTSGLNCEINFDCA 650
Qy 77 RTAYRNGE--KTYRRKSQCPCPFYEGSEMC--VPHCADK----- 112
Db 651 SNPLCHGACVDGINRYSVCSPF--TQRCNIDIDEASNPCKRDACINDVNGFRMC 708
Qy 113 -----CVHGRC--IAPNTQCCEPVGNGTNCSSACDGDHMGPHCT 149
Db 709 PEGPHHPSYQVNECLSSPCIHGNCTGTLGSLGKCLDAGVGINCE--VDKN-----ECL 762

```

```

Qy 150 SRCCKNALCNPTIGA--CHCAAGFRGMRCE---DRC-----EGTYGNDCH-QRCQC 197
Db 763 SN-PCQNGTGCTNNLVNRYCRCTCKGFKYCNQVNDIBECASNPCLNQGLDVSGLYCHC 821
Qy 198 Q---NGATCDVHTGECRCRPPYTGAFCECD-----LCPPGKHGPCE---ORC---P 239
Db 822 MLPYTKRNCQYVLAPCSINPCENNAVCKEARNEFSTCLCAPMGOGQCTYDVECVSKP 881
Qy 240 CQNGVCHHYTGE--CSCPSGMGTVCQOPCEGFRKNCQEC---QCHNGTC--DAA 292
Db 882 CMNNGICHTNGSYMCECPGFGSGMDCBEDI-----NDCLANPCQNGSCVDKVN 931
Qy 293 TGQCHCSPTGGERCODE-----CPVGYGYLC---ATGC 324
Db 932 TFSCLCLPFGVADKQYDMDNECLSEPCNKNGTCSDYNSYCTCPAGHGVCENNIDEC 991
Qy 325 ---QVNGKCYHYSG---ACLEAGFAGRC-----EARLPEGLYCIKC 364
Db 992 TESSCFNGTIC--VDGINSFCLCPVGTGFPCLHDINECSSNPCLNGLCTVDGLGTYRC 1049
Qy 365 -----DKRC-----PCHLENTSCHPMSEGCACRPGMSGLYCNE----- 398
Db 1050 TCPLGTYGKNCQTLVNLCSPPCKNKGTCAOEKARPRCLCPGMDGAYCDVILNYSCKMAA 1109
Qy 399 -----TCSPGFYGACQ---ICS---CONGACDDSVTG 426
Db 1110 LQKGVPEHLCOHSGICINAGNTHHCQPLDTGYSCEQLDEACSNPCQHGATCSDFIG 1169
Qy 427 --KCTCAGFRKIDCSFPCPLGTGTCINSSRCGCKNDVAVCSCTCKAGMHWVDCSI 484
Db 1170 GYRCECVAGCYGVNCE-----YEVDECOQNPQONGSTCIDLVNHFCS----- 1212
Qy 485 RCPSTGWFEGC--NL-TC---QCLNGAGC--NTLDG--TCTCAPGMRGEKC-----L 527
Db 1213 -CPGTBRLGLEEMIDDCAGAPHCINGCQYDRIGYSCRLPGRFAGRCEDINECLSN 1271
Qy 528 PC-ODGTGVLNCAE-----RDCDASHA-----DCCH-----PTG 555
Db 1272 PCSSEGS--LDCIDLKNNYQCVCSAFTGRHCEFTLDVCPKPCPLNGSTCAVANSVPDGF 1329
Qy 556 HCRCLPGMSGVCHCDVYCAEGRW-----GPNCSLP 584
Db 1330 ICRCPGFSGARQSSQGVVCRNGEQCVHTASGHCHCIP 1369

RESULT 11
A40701
tenascin-X precursor - human
C:Species: Homo sapiens (man)
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Dec-1999
C:Accession: A40701; A31725; C42175
R:Bristow, J.; Tee, M.K.; Gitelman, S.E.; Mellon, S.H.; Miller, W.L.
J. Cell Biol. 122, 265-278, 1993
A:Title: Tenascin-X: a novel extracellular matrix protein encoded by the human XB gene.
A:Reference number: A40701; MUID:93300909; PMID:7686164
A:Accession: A40701
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-3566 <BRI>
A:Cross-references: EMBL:X71937
R:Morel, Y.; Bristow, J.; Gitelman, S.E.; Miller, W.L.
Proc. Natl. Acad. Sci. U.S.A. 86, 6582-6586, 1989
A:Title: Transcript encoded on the opposite strand of the human steroid 21-hydroxylase gene.
A:Reference number: A31725; MUID:93367293; PMID:2475872
A:Accession: A31725
A:Molecule type: mRNA
A:Residues: 2748-3199, 'V', 3201-3298, 'E', 3299-3314, 'G', 3316-3566 <MOR>
A:Cross-references: GB:M25813; NID:g183069; PIRID:AAA35884.1; PID:g183070
R:Matsuno, K.; Arai, M.; Ishihara, N.; Ando, A.; Inoko, H.; Ikemura, T.
Genomics 12, 485-491, 1992
A:Title: Cluster of fibronectin type III repeats found in the human major histocompa
enascin.

```

A:Reference number: A42175; MUID:92217969; PMID:1373119
 A:Accession: C42175
 A:Molecule type: DNA
 A:Residues: 1849-1936 <MAT>
 A:Experimental source: clone 3.9Kf3-1
 A:Note: sequence extracted from NCBI backbone (NCBIP:95694)
 C:Genetics:
 A:Gene: GDB:TNXA: D6S103E; TNX; XA; XB
 A:Cross-references: GDB:568487; OMIM:600261
 A:Map position: 6p21.3-6p21.3
 C:Superfamily: tenascin-X; EGF homology; fibrinogen beta/gamma homology; fibronectin ty
 C:Keywords: extracellular matrix; glycoprotein
 F:435-461/Domain: EGF homology <EGF>
 F:748-828/Domain: fibronectin type III repeat homology <3F1>
 F:829-856/Domain: fibronectin type III repeat homology <3F2>
 F:873-953/Domain: fibronectin type III repeat homology <3F3>
 F:975-1055/Domain: fibronectin type III repeat homology <3F4>
 F:1078-1158/Domain: fibronectin type III repeat homology <3F5>
 F:1167-1247/Domain: fibronectin type III repeat homology <3F6>
 F:1248-1317/Domain: fibronectin type III repeat homology <3F7>
 F:1323-1403/Domain: fibronectin type III repeat homology <3F8>
 F:1412-1492/Domain: fibronectin type III repeat homology <3F9>
 F:1510-1590/Domain: fibronectin type III repeat homology <3F10>
 F:1618-1676/Domain: fibronectin type III repeat homology <3F11>
 F:1678-1749/Domain: fibronectin type III repeat homology <3F12>
 F:1751-1831/Domain: fibronectin type III repeat homology <3F13>
 F:1849-1929/Domain: fibronectin type III repeat homology <3F14>
 F:1955-2035/Domain: fibronectin type III repeat homology <3F15>
 F:2061-2141/Domain: fibronectin type III repeat homology <3F16>
 F:2167-2246/Domain: fibronectin type III repeat homology <3F17>
 F:2274-2354/Domain: fibronectin type III repeat homology <3F18>
 F:2382-2462/Domain: fibronectin type III repeat homology <3F19>
 F:2488-2568/Domain: fibronectin type III repeat homology <3F20>
 F:2584-2664/Domain: fibronectin type III repeat homology <3F21>
 F:2671-2757/Domain: fibronectin type III repeat homology <3F22>
 F:2771-2851/Domain: fibronectin type III repeat homology <3F23>
 F:2878-2958/Domain: fibronectin type III repeat homology <3F24>
 F:2977-3067/Domain: fibronectin type III repeat homology <3F25>
 F:3078-3159/Domain: fibronectin type III repeat homology <3F26>
 F:3167-3247/Domain: fibronectin type III repeat homology <3F27>
 F:3255-3334/Domain: fibronectin type III repeat homology <3F28>
 F:3349-3557/Domain: fibrinogen beta/gamma homology <F8G>
 Query Match 18.9%; Score 682; DB 1; Length 3566;
 Best Local Similarity 28.4%; Pred. No. 1.6e-30;
 Matches 191; Conservative 41; Mismatches 178; Indels 262; Gaps 38;

Db 418 RCVCWPGYGTGDDGSRACPRDCRGRCRNGVCVNAAGYSGEDCGVRSPGDCRGRCRCE 477
 Oy 423 SVTKCTCAPGFKGIDCST-----PCPLGTGYINCSS-RC--GCKND 461
 Db 478 S--GRCMKPRTYGRCCGRACPRDCRGRCRVDGRCVNCNPGTGTGDCSRRCRPGCRH 535
 Oy 462 AVCSPYDSSCTCAGNHGYDCSTR-CPSTGWFGNLTQCLANG----- 504
 Db 536 GLCE--DGVCDAGYSGEDCSTRSCPGCRGRG-----QCLDGRVCEDGYSGEDCGVR 588
 Oy 505 -----GACNTLDTCTCAPGNGRKECLP-----CODGYGLN 537
 Db 589 QCPNDCSQHGVC--DGVCLWEGYVEDCSITCTPCSNHGRCRCEGRCLDPGTGFP 646
 Oy 538 CAER---CDCSHADGCHPTTGHRCCLPGMSGVHC-----DSY 571
 Db 647 CATRMCPACRGRGR--VGVCLCHVYGGEDCGGEEPPASACPGGCPRELCTRAGQCV 704
 Oy 572 CAEGRMPNCSL 583
 Db 705 CVEGFRGPDCAI 716
 RESULT 12
 T31070
 notch homolog - sea urchin (Lytechinus variegatus)
 C:Species: Lytechinus variegatus (variegated urchin)
 C>Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 31-Jan-2000
 C:Accession: T31070
 R:Sherwood, D.R.; McClay, D.R.
 Development 124, 3363-3374, 1997
 A:Title: Identification and localization of a sea urchin Notch homologue: Insights in
 A:Reference number: Z20966; MUID:97454256; PMID:9310331
 A:Accession: T31070
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-2531 <SHG>
 A:Cross-references: EMBL:AF000634; NID:92570350; PID:92570351; PIDN:AA82088.1
 C:Superfamily: notch protein; ankyrin repeat homology; EGF homology
 Query Match 18.9%; Score 681.5; DB 2; Length 2531;
 Best Local Similarity 24.2%; Pred. No. 1.4e-30;
 Matches 216; Conservative 68; Mismatches 215; Indels 393; Gaps 53;

Db 702 SNINESSNPVNGSGCHDVNYSCECMAGYTGTRCTDPEBC-----SSNPGQ 751
 QY 328 NGKGC--YVSGACLCCEAGFAGEARL-----CPEGLYG 361
 Db 752 HGGTCNRAFAFYCTCAQAYTGTGNCENIDDCVDEPLNGIGCIDDEVNSFOCCVCPQTFVG 811
 QY 362 IKCD-KRCCHLENTSHSCHPMSEGCACRPGMSGLYCENCSGFEYGEAQO-----ICS 414
 Db 812 LICEFRSPC---EDNQCQ-NGATCYVSEDTAGYSCR--CTSGFCQNFCDDDRNECLFSP 865
 QY 415 CONGADCDSDVTG--KCTCAPGFKGIDC-----STP-----CPLGT 447
 Db 866 CRRGSGCTNLBESFECSPLGPDGPICEINIDECASGPGCTNGGICITDLIDYFCSCQNGP 925
 QY 448 YGINC---SSRC---GCKNDAYCSP-VDG-SCTCKRGMHGVDCSICPGSGTGMFCGNLTC 499
 Db 926 TERNQNDDECLSSPCRRGATCHEYVDYTCCLVGFSGMHCEINDOCT-----TS 978
 QY 500 QCLNGACNTLDG---TCTCAPGMRGECCEL-----PCDDGTGLNCAER----- 541
 Db 979 SCLYGTC--IDGVNYSYTCBCTGYTGSNCQLEINECDSDBENGA---TCDDRFSYSYC 1033
 QY 542 -CD-----CSH-ADGCHP-----TTGH--CRLPGMSGVHCD----- 569
 Db 1034 HCDVGFGLNCHRYVQWCSPONNPCYNATCYAMGHLIECHCASNWIKLCIDVPKYSIDI 1093
 QY 570 -----SYCAGRMGPNP-----SLPCY 586
 Db 1094 AASDKNVTRELCNLNGTCTIDATSSHSLCQDQGTGSGYCEVNIDECASAPCH 1145

RESULT 13 A24420

notch protein - fruit fly (Drosophila melanogaster)
 M:Alternate names: neurogenic repetitive locus protein
 C:Species: Drosophila melanogaster
 C:Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text change 10-Sep-1999
 A:Accession: A24420; A24768; S09358; A05267
 R:Kidd, S.; Kelley, M.R.; Young, M.W.
 Mol. Cell. Biol. 6, 3094-3108, 1986
 A:Reference number: A24420; MUID:87064624; PMID:3097517
 A:Accession: A24420
 A:Molecule type: DNA
 A:Residues: 1-2703 <KID>
 A:Cross-references: GB:K03508; NID:9157993; PIDN:AA28725.1; PID:9157993
 R:Wharton, K.A.; Johansen, K.M.; Xu, T.; Artavanis-Tsakonas, S.
 Cell 43, 567-581, 1985
 A:Reference number: A24768; MUID:86079539; PMID:3935325
 A:Accession: A24768
 A:Molecule type: mRNA
 A:Residues: 1-48, '1', 50-118, 'R', 120-230, 'T', 232-256, 'N', 258-266, 'A', 268-872, 'R', 874-958,
 A>Note: the authors translated the codon ATC for residue 49 as Thr, ATT for residue 2044
 R:Paulz, D.
 Nucleic Acids Res. 17, 6463-6471, 1989
 A:Title: Hypervariability of simple sequences as a general source for polymorphic DNA ma
 A:Reference number: S09358; MUID:89385974; PMID:2780284
 A:Accession: S09358
 A:Molecule type: DNA
 A:Residues: 2505-2551, 'Q'Q'Q', 2552-2576, 'E', 2578-2604 <TAU>
 R:Wharton, K.A.; Yedvobnick, B.; Flinnerty, V.G.; Artavanis-Tsakonas, S.
 Cell 40, 55-62, 1985
 A:Title: opa: a novel family of transcribed repeats shared by the Notch locus and other
 A:Reference number: A05267; MUID:85099329; PMID:2981631
 A:Accession: A05267
 A:Molecule type: DNA
 A:Residues: 2504-2576, 'E', 2578-2611 <WHA2>
 C:Genetics:
 A:Gene: notch: opa
 A:Cross-references: FlyBase:FBgn0004647
 A:Map position: 8.96-9.36
 A:Introns: 53/3; 84/3; 171/3; 240/3; 283/3; 2333/3; 2436/3; 2588/3
 C:Superfamily: notch protein; ankyrin repeat homology; EGF homology

C:Keywords: differentiation; tandem repeat; transmembrane protein
 F:27-43/Domain: transmembrane #status predicted <TM1>
 F:297-328/Domain: EGF homology <EGF1>
 F:530-561/Domain: EGF homology <EGF1>
 F:568-599/Domain: EGF homology <EGF>
 F:988-1019/Domain: EGF homology <EGX2>
 F:1064-1095/Domain: EGF homology <EGX3>
 F:1187-1218/Domain: EGF homology <EGX3>
 F:1746-1762/Domain: transmembrane #status predicted <TM2>
 F:1950-1982/Domain: ankyrin repeat homology <AN1>
 F:1988-2015/Domain: ankyrin repeat homology <AN2>
 F:1988-2044/Domain: transmembrane #status predicted <TM3>
 F:2017-2049/Domain: ankyrin repeat homology <AN3>
 F:2050-2082/Domain: ankyrin repeat homology <AN4>
 F:2088-2115/Domain: ankyrin repeat homology <AN5>
 F:2538-2568/Region: glutamine-rich
 F:2538-2568/Domain: neurogenic repetitive element #status predicted <OPA>
 Query Match 18.8%; Score 677; DB 1; Length 2703;
 Best Local Similarity 25.4%; Pore No. 2.6e-30;
 Matches 208; Conservative 78; Mismatches 203; Indels 330; Gaps 51;
 QY 7 SCL---SFICLLCHWICTASPLNED--PNVCSHESYSVTVDGSHPHPDQIYITSC 60
 Db 502 SCLDPEGTFRVCMFGFTGCEIDIDECQSNPC-----LNDGTC 541
 QY 61 TDLINFKCTRRHVSRTAIRRIGETMYRRKSCCCGFYESGMC--YPRCADKVVHR 117
 Db 542 HDKINFGKS-----CALGF--TGARCOINIDDCQSQPCNRN 576
 QY 118 -----CIAPNFCOCEPMGNTGNS--SACDGDHMGPHCTSRQCNGALCNPTG-ACH 168
 Db 577 GICHDSIAGYSCCEPRPGYGTGSEININDCSN-----PCIRGKCIDVNSFKKL 626
 QY 169 CAAGFRGW-----RCEDR-----CEGTYG-----NDCHQRCQ 196
 Db 627 CBPGYGYICOKQKQINCEESNPQCFDGHODRVGSYYCQCAQSGRKNCEVNECHSN-P 685
 QY 197 CONGATC-DHYMG-ECRCPRGYTGARFCELCPRGKRGPCQEQRCPCONGVC-HYNTG-E 252
 Db 686 CNGGATCIDGINSYKQCVPGFTGHCE-----KNVDKIS-SPANNGVCIDOVNGYK 738
 QY 253 GSCPSGMGTVC-----GOP-----CEGRGRKNCS--OECQ-- 282
 Db 739 CCEPRGFYAHCLSDVDECAASNPCVNEGRCEDEINEFICHPRGYKXKCEIDIDECSSN 798
 QY 283 -CHNGTC-DAATG-QCHCSPGYTGERQ--DECPVGYGYVCAFTCCQVNGKCY-HV 335
 Db 799 PCOHGCTCYDKLNAFSCQCPMGYTGQKCEINIDDC-----YTNPCCNGGTCIDIV 848
 QY 336 SG-ACICEAGFPAGERCEARLCEPGLGICR-DKRCCHLENTSHCHPMNSG--ECACKP 389
 Db 849 NGYKCKCKVPFTGRDCEKMDP-----CASNRK--KNEAKTPSSNFIIDFCTCTKL 897
 QY 390 GMSGLYCNE-----TCSPGFYGEAC--QOICS--CON 417
 Db 898 GTTGRUCDDIDECSSISPCRRNGASCLNPGSYRCLCTGTGEBRDCAINMTDCASPCQN 957
 QY 418 GADCDSDVTG--KCTCAPGFKGIDCST-----PCLCTYGI 450
 Db 958 GGTGIDIGIDYDCLCDVDPGKHCEFDINECLSQPCONGATCSQVYNSYTCPCPLGFSGI 1017
 QY 451 NCS-----SRGCKRDVANCSPVDG--SCNCKAGMHGVDSIR----- 485
 Db 1018 NCQTNDECTESSCLNNGSC--IDGINGYNSCLAGYSGANCOYKLNKCDNSNPLCNGATC 1075
 QY 486 -----CPSGTWGFECNLT-----TCQCLNGACNTL--DGTCTCAPGMRGECPC- 526
 Db 1076 HGOONNRYTCHCPSGFTGKCKSEYVDMCGSPCENAGTCSQMKHOPCKCKSAGMTGLCDV 1135
 QY 527 --LPCODGT--YGLNCAERCD--CSHADGCHPTTGHCRCLPGMSGVHC----- 568
 Db 1136 QRTSQDADADRGLSLRQLCNGNGTCKDYGNSHV-----CYCSGYAGSYCQKEIDECQSQP 1191

OY 569 -----DSVCAEGRMGPNCSL-----PC 585
 Db 1192 CONGTCDLIGATGECQCRQGFQGNCLINDCAPNFC 1230

RESULT 14 T09070

probable tenascin X - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jan-2000
 C:Accession: T09070
 R:Rowen, L.; Mahairas, G.; Qin, S.; Ahearn, M.E.; Dankers, C.; Lasky, S.; Loretz, C.; Se
 submitted to the EMBL Data Library, October 1997
 A:Description: Sequence of the mouse major histocompatibility locus class III region.
 A:Reference number: 216543
 A:Accession: T09070
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-4006 <ROM>
 A:Cross-references: EMBL:AF030001; NID:g2564945; PID:g2564958
 C:Genetics:
 A:Gene: TNX
 A:Map position: 17
 A:Intons: 124/1; 735/1; 826/1; 914/1; 1037/1; 1135/1; 1232/1; 1329/1; 1440/1; 15
 019/1; 3115/1; 3208/1; 3302/1; 3405/1; 3517/1; 3558/1; 3606/1; 3646/1; 3694/1; 3737/3;
 C:Superfamily: tenascin-X; EGF homology; fibrinogen beta/gamma homology; fibronectin ty
 F:432-448/Domain: EGF homology <EGF>
 F:826-906/Domain: fibronectin type III repeat homology <3PR>
 F:3789-3997/Domain: fibrinogen beta/gamma homology <F8g>

Query Match 18.88; Score 676.5; DB 2; Length 4006;
 Best Local Similarity 29.18; Pred. No. 3.5e-30;

Matches 190; Conservative 50; Mismatches 178; Indels 235; Gaps 38;

OY 82 HGEKTRRKSSQCCPGF-----YESGEMCVPHCAD--KCYHGRCIAPNT 123
 Db 135 HGVDLSKACSCPEGMGPTCDPTDTKTSSPSKSCDECDNDRGRCYRCYV----- 190
 OY 124 CQCEPBGNGTNC--SACDGDHMGHCTSRCKCKNGALCNPTTGACHCAAGFRGMRCEDRC 182
 Db 191 --CPFGTSGPSCWSPSCPGD-----CQGRGRC-----YKGVCAVCAAGSG--PDC 231
 OY 183 EGVGNDCHQRCOCQNGATCDHYTGECRCPGYTGAF--EDLPPGKHGQCQRCCQ 241
 Db 232 SQRCSPRNCNDRGRC--GRCVCDPRTSGEDCGVRSRPG-----CSQGRGRC 278
 OY 242 NGVCHHTVECSQPSGMGTVC--QCPREGFRGKNSQECQHNHGTCDAAATGQCHSP 300
 Db 279 N-----GLCVNPGVSGEDCGVRNCPRG-----CSQGRGRC-----GRCVCDP 317
 OY 301 GYTGRCODECPVGTGYLCAETC--OCVNGKCYHVSAGACLCAAGFAGERCEARLCPREG 358
 Db 318 GYSESDS-----MRTCPMDCGDGGRG--VDGRKVCWPGVSGEDCSRTTGRPD 363
 OY 359 LYG1-KC-DKRCPCHL-----NTHSC-----HPMSECCCKGMSLYCNE----- 398
 Db 364 CRGRGRDEDECTDAGYSSDDCGVRSCPGDCNDRGCHCEGRCCWPGYTGADSTNACP 423
 OY 399 -----TCSPGFYGCAC--QOIC--SCQNGADDSYVGTCTCAAFGKIDCST 441
 Db 424 RDCRGRCEDGVCAVCHAGVSGEDCGVRSRPGDRGNGNES--GRCVCPGYTGRCCT 481
 OY 442 -----PPLGTGYNCS--RC--GCKADAVCSPYDSCTCTKAGMHV 480
 Db 482 RACPGDCRGRCVDRGCVNCPFTGEDCGSRRCPGDRGHGCE--NGVCAVAVGSGD 539
 OY 481 DCSIR--CPSGTWGFGCMLTQCLNG-----GACTTLDGTC 514
 Db 540 DCSIRSPSCRGRC-----QCLNGLCEDDEGVSGEDCGTRCPDRCSQHGVCQ--DGLC 592
 OY 515 TCAPGMGKCEKEL-----PCODGTGYN-----CAER--CDCSHADGCHPT 553

Db 593 MCHAGYAGEDCSIRTCPADCRRRGRCEGRVCNCGYTGPAATTCPADCRGRC--V 650
 OY 554 TGHCRCLPQMSGVHC-----DSVCAEGRMGPNCSL 583
 Db 651 QGVCMCYGYSGEDCGQEBRPASACPGCGPRELCRAQGVCEVGRFPDCAI 703

RESULT 15 S18188

notch protein homolog - rat
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 19-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 02-Aug-2002
 C:Accession: S18188
 R:Weinmaster, G.; Roberts, V.J.; Lemke, G.
 Development 113, 199-205, 1991
 A:Title: A homolog of Drosophila Notch expressed during mammalian development.
 A:Reference number: S18188; MUID:92111383; PMID:1764995
 A:Accession: S18188
 A:Molecule type: mRNA
 A:Residues: 1-2531 <WEI>
 A:Cross-references: EMBL:X57405; NID:g57634; PID:g57635
 C:Superfamily: notch protein; ankyrin repeat homology; EGF homology
 F:987-1018/Domain: EGF homology <EGF1>
 F:1025-1056/Domain: EGF homology <EGF2>
 F:1233-1264/Domain: EGF homology <EGF3>
 F:1917-1949/Domain: ankyrin repeat homology <AN1>
 F:1950-1982/Domain: ankyrin repeat homology <AN2>
 F:1984-2016/Domain: ankyrin repeat homology <AN3>
 F:2017-2049/Domain: ankyrin repeat homology <AN4>
 F:2050-2082/Domain: ankyrin repeat homology <AN5>

Query Match 18.78; Score 675; DB 2; Length 2531;
 Best Local Similarity 25.78; Pred. No. 3.2e-30;
 Matches 208; Conservative 70; Mismatches 206; Indels 324; Gaps 48;

OY 4 S1NCSLS-----FICLLCHWIGTASPLNEDPNVCSHWEISVTVQESY 48
 Db 603 NINECHSQPCRHHGTCQDRDNYVILCLKGTTPCEINLID--CA----- 646
 OY 49 PHPDQIYVTSCTDILNMFKCTRRHVSRYTAYRHGKTYRRKSQCCPGFESGPMC--- 105
 Db 647 SNPCDS--GTCIDKIDYECA-----CEGTY--TSMCAVN 678
 OY 106 VPHCA-----DKCVHGRG--IAPNT 123
 Db 679 IDECAGSPCHNGTCEDEGIAGFTCRCEGYHDPCLSEVNECNPNCTHGCACRGLNGYK 738
 OY 124 CQCEPBGNGTNC--SSACDGDHMGHCTSRCKCKNGALCNPTTG--ACHCAAGFRGMRC 178
 Db 739 CDCAPGMSGTNCIDINNCESEN-----PCVNGGTCKDMTSGYVCTCRGFGSPNC 788
 OY 179 EDRCEQTYGNDCHQRCQCGNGATC--DHYTG--ECRCPPGYTGAFCEBLCPPGKHGPOCEQ 236
 Db 789 Q-----TINNEGASN--PCLNQGTCTIDVAGYKNCCLPPTTGACVAYLAP-----C-A 834
 OY 237 RCPQNGVC--HAYTGECSPSGMGTVC-----GQCPREGR----- 272
 Db 835 TSPCKNSGVKESDEYSEFCVCPMGQGTCEIDINECVKSPCRHGASCONGYSRCL 894
 OY 273 -----FGKNS--QEQ-----CHNGTCDAAATG--CHCSPGTGECODE----- 310
 Db 895 CQAGYTRNCESDIDRPNPCHNGSGCTDGVNAFAFDCLPFGGACAEEDINECATNPC 954
 OY 311 -----CPVGTGYLCAET-----CQCVNGKCYHVS-----ACLCGAG 344
 Db 955 QNAGNCTDVCDSYTCPTPFENGHCENNPDPCTESSCFNGTIC--VDGINSFCLCPPG 1012
 OY 345 PABRCEARLCPGLYGI--KCDKRCPCHLNTHSCHPMG--DCACKPGMSGLYCNE--- 398
 Db 1013 FTGSGYQ-----YDVNCDISR--PCLHGGT--CQDSYGYKCTCPGGTGLMCQQLVR 1061
 OY 399 --TCSPGFYGCACQI-----CSCQN--GADDSYTGKCTCAPRFGGIDCSTPCPLGTY 448


```

Db 1062 WCDAPCKNGKCMQNTMTQYHCEBRSWGTGNCVLSYSCVFAAQRGIDVTLLCQHGL 1121
QY 449 GIN-----CSSRCG-----CKNDAYCSPVDG--SCTKAGWHGYDCS 483
Db 1122 CVDEEDRHHYCHQAGYGYSCYCEDEVEDCSPNPNCONGATCFDYLGGFSCKCVAGYHGSNCS 1181
QY 484 -----IRPSTGTMGFCNLTF--C-----QCLUNG 504
Db 1182 BEINECLSQPCQNGGTCTIDLTNTYKSCPRGTGCVHCEINVDCHPPLDPASRSPKCFNN 1241
QY 505 GACNTLIDG--TCTCAPGMGEKCE-----LPCOD-GTYGLNCAERKDCSHADGCHPT 553
Db 1242 GTCVDQYGGYTCICPPGFVGBERCEGDVNECLSNPCDPRGTQ--NCVQRYN----- 1289
QY 554 TGHCRCLPGWSGVHCDVSCAEGRWGPNC 581
Db 1290 DFHCECRAGHTGRRCESV-INGCRGKPC 1316

```

Search completed: May 9, 2003, 11:58:44
 Job time : 34.9757 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 9, 2003, 10:55:56 ; Search time 9.50637 Seconds

(Without alignments)
2556.719 Million cell updates/sec

File: US-10-092-390-4
Database: 3601
Sequence: 1 MVLINSLCFLICLCHWL.....HCDVCAEGRWPCSLPCY 586

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database: SWISSPROT_40

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	843.5	23.4	1 SREC_HUMAN	Q14162 homo sapien
2	719	20.0	1 NOTC_XENLA	P21783 xenopus lae
3	717	19.9	1 NOTC1_BRARE	P46530 brachydanio
4	697	19.4	1 NOTC3_HUMAN	Q9um47 homo sapien
5	693	19.2	1 NOTC1_MOUSE	O01705 mus musculu
6	685.5	19.0	1 NOTC3_MOUSE	O61982 mus musculu
7	685.5	19.0	1 NOTC2_RAT	Q9qW30 mus musculu
8	682.5	19.0	1 NOTC3_RAT	Q9q172 ratius norv
9	682	18.9	1 TENX_HUMAN	Q91172 ratius norv
10	677	18.8	1 NOTC2_MOUSE	O35516 homo sapien
11	677	18.8	1 NOTC2_MOUSE	P07207 drosophila
12	675.5	18.8	1 NOTC4_HUMAN	Q99466 homo sapien
13	675	18.5	1 NOTC1_RAT	Q07008 ratius norv
14	667.5	18.5	1 JAG3_BRARE	Q90Y54 brachydanio
15	666.5	18.5	1 NOTC2_HUMAN	Q04721 homo sapien
16	664.5	18.5	1 FBPI_STRPU	P10079 strongyloce
17	662	18.4	1 NOTC1_HUMAN	P16531 homo sapien
18	658	18.3	1 NOTC4_MOUSE	P11693 mus musculu
19	646	17.9	1 TENA_CHICK	P10033 gallus galli
20	644.5	17.9	1 TENA_HUMAN	P14821 homo sapien
21	638	17.7	1 LMA5_HUMAN	O15230 homo sapien
22	633	17.6	1 JAG2_HUMAN	O9Y212 homo sapien
23	631.5	17.5	1 CRB_DROME	P10040 drosophila
24	631	17.5	1 JAG2_MOUSE	O9qy65 mus musculu
25	629	17.5	1 JAG1_HUMAN	P78504 homo sapien
26	619.5	17.2	1 JAG2_RAT	O96722 ratius norv
27	616	17.1	1 JAG1_RAT	O6Y057 brachydanio
28	615	17.1	1 JAG1_BRARE	O9qxx1 mus musculu
29	614	17.1	1 JAG1_MOUSE	O29116 sus scrofa
30	611	17.0	1 TENA_PIG	Q21313 caenorhabdi
31	593.5	16.5	1 LML2_CAEBL	P18168 drosophila
32	587	16.3	1 SERR_DROME	O61001 mus musculu
33	586	16.3	1 LMA5_MOUSE	

ALIGNMENTS

RESULT 1	SREC_HUMAN	STANDARD	PRT	830 AA.
AC	Q14162; 043701;			
DT	15-JUN-2002 (Rel. 41, Created)			
DT	15-JUN-2002 (Rel. 41, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Endothelial cells scavenger receptor precursor (Acetyl LDL receptor).			
GN	SREC OR KIAA0149.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
OX	NCBI_Taxid:9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Umbilical vein endothelial cells;			
KX	MEDLINE=9805897; PubMed=9395444;			
RA	Adachi H., Tsujimoto M., Arai H., Inoue K.;			
RT	"Expression cloning of a novel scavenger receptor from human			
RT	endothelial cells."			
RL	J. Biol. Chem. 272:31217-31220(1997).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Bone marrow;			
KX	MEDLINE=96127530; PubMed=8590280;			
RA	Nagase T., Seki N., Tanaka A., Ishikawa K.-I., Nomura N.;			
RT	"Prediction of the coding sequences of unidentified human genes. IV.			
RT	The coding sequences of 40 new genes (KIAA0121-KIAA0160) deduced by			
RT	analysis of cDNA clones from human cell line KG-1."			
RL	DNA Res. 2:167-174(1995).			
CC	-1- FUNCTION: Mediates the binding and degradation of acetylated low			
CC	density lipoprotein (Ac-LDL).			
CC	-1- SUBCELLULAR LOCATION: Type I membrane protein (Potential).			
CC	-1- TISSUE SPECIFICITY: Endothelial cells.			
CC	-1- SIMILARITY: CONTAINS 6 EGF-LIKE DOMAINS.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@sib-sib.ch).			
CC	-----			
DR	EMBL; D66864; BAA24070.1; -			
DR	EMBL; D63483; BAA09770.1; -			
DR	HSSP; P01180; 2B2.			
DR	InterPro: IPR000561; EGF-like.			
DR	PROSITE; PS00022; EGF_1; 6.			
KW	Receptor; Glycoprotein; Signal; Transmembrane; EGF-like domain;			
KW	Repeat.			
FT	SIGNAL	1	19	POTENTIAL.
FT	CHAIN	20	830	ENDOTHELIAL CELLS SCAVENGER RECEPTOR.
FT	DOMAIN	20	421	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	422	442	POTENTIAL.

```

FT DOMAIN 443 830 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 53 87 EGF-LIKE 1.
FT DOMAIN 95 130 EGF-LIKE 2.
FT DOMAIN 155 191 EGF-LIKE 3.
FT DOMAIN 215 249 EGF-LIKE 4.
FT DOMAIN 302 339 EGF-LIKE 5.
FT DOMAIN 351 382 EGF-LIKE 6.
FT DOMAIN 476 620 PRO/SER-RICH.
FT DOMAIN 622 798 GLY-RICH.
FT DOMAIN 431 438 POLY-LEU.
FT DISULFID 57 69 POTENTIAL.
FT DISULFID 63 75 POTENTIAL.
FT DISULFID 77 86 POTENTIAL.
FT DISULFID 99 111 POTENTIAL.
FT DISULFID 105 118 POTENTIAL.
FT DISULFID 120 129 POTENTIAL.
FT DISULFID 159 172 POTENTIAL.
FT DISULFID 165 179 POTENTIAL.
FT DISULFID 181 190 POTENTIAL.
FT DISULFID 219 230 POTENTIAL.
FT DISULFID 225 237 POTENTIAL.
FT DISULFID 239 248 POTENTIAL.
FT DISULFID 306 319 POTENTIAL.
FT DISULFID 313 326 POTENTIAL.
FT DISULFID 329 338 POTENTIAL.
FT DISULFID 353 363 POTENTIAL.
FT DISULFID 372 381 POTENTIAL.
FT CAROHND 289 289 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CAROHND 382 382 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CAROHND 393 393 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 662 662 R -> W (IN REF. 2).
SQ SEQUENCE 830 AA; 87430 MW; F560D9E1A64D779 CRC64;

```

```

Query Match 23.4%; Score 843.5; DB 1; Length 830;
Best Local Similarity 36.2%; Pred. No. 7.6e-47;
Matches 158; Conservative 51; Mismatches 164; Indels 63; Gaps 15;

```

```

QY 93 OCCPGFSEGEHC-VPHC--ADKCYHGR-CIAPNTQCCEPFGMGTCSSACDGDHWHPHC 148
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 40 OCCACMRQKRODECTIPICGPRACQKDEVCVRPGLRCRKFEGFAGSSRCRPGQYWGPD 99
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 149 TSRCCCKNALNPITGACHCAAGFRGWCEDRCBEGTIGNDCHQCCQONATGCHYVG 208
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 100 RESCPHGHQCEPATGACQCADRWGARCERPCACGPHGR-----CDPATG 146
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 209 ECRCPFGYTGACEDLCPPGKHGPOCEORPCQNGGVCHHTGECSPGMMGTVCQGPC 268
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 147 VCHCEGGMSSSTCRRCQCCNTAAREQ-----ATGACVCRKGM----- 185
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 269 PGRGFRGKNSQRCQCHNGSTCAATGQCHCSFGYTGERCQDECPVGTGYVLAETGQCVN 328
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 186 ---WGRRCSPFRNCCH-GSPCEQDSRCACRPGMWCPECCQO-----CECYR 227
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 329 GSKCYHVSACLCCEAGFNGERCEARLCPGLGKICDKRCPHLENTHSCHPMSGEC-AC 387
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 228 -GRCSAASECTCPGFRARCELP-CPAGSHGVQCAHSCG-RCKINERCSPTGSCESC 284
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 388 KPGMSGLYCNETCSPGFYEGACQOIC-SCONGADCDVYFK-TCAPGFKGIDSPPCPL 445
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 285 EFGMNGTQCOQCLPCTGTFEESCEQCPKRRHGEACEPTDGHORCDPGLGRCCEPCT 344
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 446 GTYGINCSSRCCKNDVAVSPYDGSCTCAAGHGVDCSIRCPSGTWGFGCNTLTCQCLNG 505
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 345 GTFGECDGCTPTVOGSCDVTGDCVCSAGYWGSPSCNMSCPAGFGNNCSVPCEPE-G 403
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 506 ACNTLDGTCTCAPGR 521
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 404 LCHPVSGSCQPGSGSR 419
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 2
NOTC_XENLA

```

ID NOTC_XENLA STANDARD: PRT; 2524 AA.
AC P21783;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Neurogenic locus notch protein homolog precursor (XOTCH protein).
GN XOTCH.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae;
OC Xenopodinae; Xenopus.
OX NCBI_Taxid=8355;
[1]
RX MEDLINE=90385285; PubMed=2402639;
RX Coffman C., Harris W., Kintner C.;
RT "Xotch, the Xenopus homolog of Drosophila notch.";
RL Science 249:1438-1441(1990).
[2]
RN REVISIONS TO 1759-1782.
RP Kintner C.;
RA Submitted (JUN-1996) to the EMBL/Genbank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED ALMOST UNIFORMLY IN EARLY EMBRYOS.
CC -1- SIMILARITY: BELONGS TO THE NOTCH FAMILY.
CC -1- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC -1- SIMILARITY: CONTAINS 6 ANK REPEATS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
-----
CC EMBL; M33874; AB02039.1; -
CC PIR; A35844; A35844.
DR HSSP; P00740; 1EDM.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000561; EGF-Like.
DR InterPro; IPR000742; EGF-2.
DR InterPro; IPR001811; EGF-Ca.
DR InterPro; IPR001438; EGF-II.
DR InterPro; IPR000800; Notch.
DR Pfam; PF00008; EGF; 36.
DR Pfam; PF00023; ank; 6.
DR Pfam; PF00066; notch; 3.
DR PRINTS; PR00010; EGFBL00D.
DR PRINTS; PR01452; NOTCH.
DR SMART; SM00248; ANK; 5.
DR SMART; SM00179; EGF_CA; 23.
DR SMART; SM00001; EGF_Like; 11.
DR SMART; SM00004; NL; 2.
DR PROSITE; PS50088; ANK_REPEAT; 4.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; 23.
DR PROSITE; PS00022; EGF_1; 34.
DR PROSITE; PS01186; EGF_2; 29.
DR PROSITE; PS01187; EGF_CA; 21.
KW Differentiation; Neurogenesis; Repeat; ANK repeat; EGF-like domain;
KW Transmembrane; Signal; Glycoprotein.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 2524 NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG.
FT DOMAIN 20 1728 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1729 1750 POTENTIAL.
FT DOMAIN 1751 2524 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 20 57 EGF-LIKE 1.
FT DOMAIN 58 99 EGF-LIKE 2.
FT DOMAIN 102 140 EGF-LIKE 3.
FT DOMAIN 141 177 EGF-LIKE 4.

```

```

FT DOMAIN 179 215 EGF-LIKE 5, CALCIIUM-BINDING (POTENTIAL).
FT DOMAIN 217 254 EGF-LIKE 6.
FT DOMAIN 256 292 EGF-LIKE 7, CALCIIUM-BINDING (POTENTIAL).
FT DOMAIN 284 332 EGF-LIKE 8, CALCIIUM-BINDING (POTENTIAL).
FT DOMAIN 334 370 EGF-LIKE 9, CALCIIUM-BINDING (POTENTIAL).
FT DOMAIN 371 409 EGF-LIKE 10.
FT DOMAIN 411 449 EGF-LIKE 11, CALCIIUM-BINDING (POTENTIAL).
FT DOMAIN 451 487 EGF-LIKE 12, CALCIIUM-BINDING (POTENTIAL).
FT DOMAIN 489 525 EGF-LIKE 13, CALCIIUM-BINDING (POTENTIAL).
FT DOMAIN 527 563 EGF-LIKE 14, CALCIIUM-BINDING (POTENTIAL).
FT DOMAIN 565 600 EGF-LIKE 15, CALCIIUM-BINDING (POTENTIAL).
FT DOMAIN 602 638 EGF-LIKE 16, CALCIIUM-BINDING (POTENTIAL).
FT DOMAIN 640 675 EGF-LIKE 17.
FT DOMAIN 677 713 EGF-LIKE 18, CALCIIUM-BINDING (POTENTIAL).
FT DOMAIN 715 750 EGF-LIKE 19, CALCIIUM-BINDING (POTENTIAL).
FT DOMAIN 752 788 EGF-LIKE 20, CALCIIUM-BINDING (POTENTIAL).
FT DOMAIN 790 826 EGF-LIKE 21, CALCIIUM-BINDING (POTENTIAL).
FT DOMAIN 828 866 EGF-LIKE 22.
FT DOMAIN 904 942 EGF-LIKE 23, CALCIIUM-BINDING (POTENTIAL).
FT DOMAIN 906 942 EGF-LIKE 24, CALCIIUM-BINDING (POTENTIAL).
FT DOMAIN 944 980 EGF-LIKE 25.
FT DOMAIN 982 1018 EGF-LIKE 26.
FT DOMAIN 1020 1056 EGF-LIKE 27, CALCIIUM-BINDING (POTENTIAL).
FT DOMAIN 1058 1094 EGF-LIKE 28.
FT DOMAIN 1096 1142 EGF-LIKE 29.
FT DOMAIN 1144 1180 EGF-LIKE 30.
FT DOMAIN 1182 1218 EGF-LIKE 31, CALCIIUM-BINDING (POTENTIAL).
FT DOMAIN 1220 1264 EGF-LIKE 32, CALCIIUM-BINDING (POTENTIAL).
FT DOMAIN 1266 1304 EGF-LIKE 33.
FT DOMAIN 1306 1346 EGF-LIKE 34.
FT DOMAIN 1347 1383 EGF-LIKE 35.
FT DOMAIN 1386 1424 EGF-LIKE 36.
FT DOMAIN 1441 1478 LIN/NORCH 1.
FT REPEAT 1479 1520 LIN/NORCH 2.
FT REPEAT 1521 1560 LIN/NORCH 3.
FT REPEAT 1561 1600 ANK 1.
FT REPEAT 1601 1640 ANK 2.
FT REPEAT 1641 1680 ANK 3.
FT REPEAT 1681 1720 ANK 4.
FT REPEAT 1721 1760 ANK 5.
FT REPEAT 1761 1800 ANK 6.
FT DISULFID 1801 1840 BY SIMILARITY.
FT DISULFID 1841 1880 BY SIMILARITY.
FT DISULFID 1881 1920 BY SIMILARITY.
FT DISULFID 1921 1960 BY SIMILARITY.
FT DISULFID 1961 2000 BY SIMILARITY.
FT DISULFID 2001 2040 BY SIMILARITY.
FT DISULFID 2041 2080 BY SIMILARITY.
FT DISULFID 2081 2120 BY SIMILARITY.
FT DISULFID 2121 2160 BY SIMILARITY.
FT DISULFID 2161 2200 BY SIMILARITY.
FT DISULFID 2201 2240 BY SIMILARITY.
FT DISULFID 2241 2280 BY SIMILARITY.
FT DISULFID 2281 2320 BY SIMILARITY.
FT DISULFID 2321 2360 BY SIMILARITY.
FT DISULFID 2361 2400 BY SIMILARITY.
FT DISULFID 2401 2440 BY SIMILARITY.
FT DISULFID 2441 2480 BY SIMILARITY.
FT DISULFID 2481 2520 BY SIMILARITY.
FT DISULFID 2521 2560 BY SIMILARITY.
FT DISULFID 2561 2600 BY SIMILARITY.
FT DISULFID 2601 2640 BY SIMILARITY.
FT DISULFID 2641 2680 BY SIMILARITY.
FT DISULFID 2681 2720 BY SIMILARITY.
FT DISULFID 2721 2760 BY SIMILARITY.
FT DISULFID 2761 2800 BY SIMILARITY.
FT DISULFID 2801 2840 BY SIMILARITY.
FT DISULFID 2841 2880 BY SIMILARITY.
FT DISULFID 2881 2920 BY SIMILARITY.
FT DISULFID 2921 2960 BY SIMILARITY.
FT DISULFID 2961 3000 BY SIMILARITY.
FT DISULFID 3001 3040 BY SIMILARITY.
FT DISULFID 3041 3080 BY SIMILARITY.
FT DISULFID 3081 3120 BY SIMILARITY.
FT DISULFID 3121 3160 BY SIMILARITY.
FT DISULFID 3161 3200 BY SIMILARITY.
FT DISULFID 3201 3240 BY SIMILARITY.
FT DISULFID 3241 3280 BY SIMILARITY.
FT DISULFID 3281 3320 BY SIMILARITY.
FT DISULFID 3321 3360 BY SIMILARITY.
FT DISULFID 3361 3400 BY SIMILARITY.
FT DISULFID 3401 3440 BY SIMILARITY.
FT DISULFID 3441 3480 BY SIMILARITY.
FT DISULFID 3481 3520 BY SIMILARITY.
FT DISULFID 3521 3560 BY SIMILARITY.
FT DISULFID 3561 3600 BY SIMILARITY.
FT DISULFID 3601 3640 BY SIMILARITY.
FT DISULFID 3641 3680 BY SIMILARITY.
FT DISULFID 3681 3720 BY SIMILARITY.
FT DISULFID 3721 3760 BY SIMILARITY.
FT DISULFID 3761 3800 BY SIMILARITY.
FT DISULFID 3801 3840 BY SIMILARITY.
FT DISULFID 3841 3880 BY SIMILARITY.
FT DISULFID 3881 3920 BY SIMILARITY.
FT DISULFID 3921 3960 BY SIMILARITY.
FT DISULFID 3961 4000 BY SIMILARITY.
FT DISULFID 4001 4040 BY SIMILARITY.
FT DISULFID 4041 4080 BY SIMILARITY.
FT DISULFID 4081 4120 BY SIMILARITY.
FT DISULFID 4121 4160 BY SIMILARITY.
FT DISULFID 4161 4200 BY SIMILARITY.
FT DISULFID 4201 4240 BY SIMILARITY.
FT DISULFID 4241 4280 BY SIMILARITY.
FT DISULFID 4281 4320 BY SIMILARITY.
FT DISULFID 4321 4360 BY SIMILARITY.
FT DISULFID 4361 4400 BY SIMILARITY.
FT DISULFID 4401 4440 BY SIMILARITY.

```

```

FT DISULFID 439 448 BY SIMILARITY.
FT DISULFID 449 458 BY SIMILARITY.
FT DISULFID 459 468 BY SIMILARITY.
FT DISULFID 469 478 BY SIMILARITY.
FT DISULFID 479 488 BY SIMILARITY.
FT DISULFID 489 498 BY SIMILARITY.
FT DISULFID 499 508 BY SIMILARITY.
FT DISULFID 509 518 BY SIMILARITY.
FT DISULFID 519 528 BY SIMILARITY.
FT DISULFID 529 538 BY SIMILARITY.
FT DISULFID 539 548 BY SIMILARITY.
FT DISULFID 549 558 BY SIMILARITY.
FT DISULFID 559 568 BY SIMILARITY.
FT DISULFID 569 578 BY SIMILARITY.
FT DISULFID 579 588 BY SIMILARITY.
FT DISULFID 589 598 BY SIMILARITY.
FT DISULFID 599 608 BY SIMILARITY.
FT DISULFID 609 618 BY SIMILARITY.
FT DISULFID 619 628 BY SIMILARITY.
FT DISULFID 629 638 BY SIMILARITY.
FT DISULFID 639 648 BY SIMILARITY.
FT DISULFID 649 658 BY SIMILARITY.
FT DISULFID 659 668 BY SIMILARITY.
FT DISULFID 669 678 BY SIMILARITY.
FT DISULFID 679 688 BY SIMILARITY.
FT DISULFID 689 698 BY SIMILARITY.
FT DISULFID 699 708 BY SIMILARITY.
FT DISULFID 709 718 BY SIMILARITY.
FT DISULFID 719 728 BY SIMILARITY.
FT DISULFID 729 738 BY SIMILARITY.
FT DISULFID 739 748 BY SIMILARITY.
FT DISULFID 749 758 BY SIMILARITY.
FT DISULFID 759 768 BY SIMILARITY.
FT DISULFID 769 778 BY SIMILARITY.
FT DISULFID 779 788 BY SIMILARITY.
FT DISULFID 789 798 BY SIMILARITY.
FT DISULFID 799 808 BY SIMILARITY.
FT DISULFID 809 818 BY SIMILARITY.
FT DISULFID 819 828 BY SIMILARITY.
FT DISULFID 829 838 BY SIMILARITY.
FT DISULFID 839 848 BY SIMILARITY.
FT DISULFID 849 858 BY SIMILARITY.
FT DISULFID 859 868 BY SIMILARITY.
FT DISULFID 869 878 BY SIMILARITY.
FT DISULFID 879 888 BY SIMILARITY.
FT DISULFID 889 898 BY SIMILARITY.
FT DISULFID 899 908 BY SIMILARITY.
FT DISULFID 909 918 BY SIMILARITY.
FT DISULFID 919 928 BY SIMILARITY.
FT DISULFID 929 938 BY SIMILARITY.
FT DISULFID 939 948 BY SIMILARITY.
FT DISULFID 949 958 BY SIMILARITY.
FT DISULFID 959 968 BY SIMILARITY.
FT DISULFID 969 978 BY SIMILARITY.
FT DISULFID 979 988 BY SIMILARITY.
FT DISULFID 989 998 BY SIMILARITY.
FT DISULFID 999 1008 BY SIMILARITY.
FT DISULFID 1009 1018 BY SIMILARITY.
FT DISULFID 1019 1028 BY SIMILARITY.
FT DISULFID 1029 1038 BY SIMILARITY.
FT DISULFID 1039 1048 BY SIMILARITY.
FT DISULFID 1049 1058 BY SIMILARITY.
FT DISULFID 1059 1068 BY SIMILARITY.
FT DISULFID 1069 1078 BY SIMILARITY.
FT DISULFID 1079 1088 BY SIMILARITY.
FT DISULFID 1089 1098 BY SIMILARITY.
FT DISULFID 1099 1108 BY SIMILARITY.
FT DISULFID 1109 1118 BY SIMILARITY.
FT DISULFID 1119 1128 BY SIMILARITY.
FT DISULFID 1129 1138 BY SIMILARITY.
FT DISULFID 1139 1148 BY SIMILARITY.
FT DISULFID 1149 1158 BY SIMILARITY.
FT DISULFID 1159 1168 BY SIMILARITY.

```

Query Match 20.0%: Score 719; DB 1; Length 2524;
 Best Local Similarity 25.8%; Pred No. 1.3e-38;
 Matches 222; Conservative 60; Mismatches 222; Indels 355; Gaps 50;

```

QY 5 INSCISFICL-----LICHMIGTASPLNED-----PNCISHM-----ESYS 41
DB 603 INECLSKKCLANGGQCTDRENGYICPCPKGTGVNCEKIDDCASMLCNGKCIDKIDGVE 662
QY 42 VYQESYRPHFQIYTT-----SCDILMWFKCTRRVRYRRAIRGERTWYRR 90
DB 663 CYCEPGYTGKLCININEDCSNPCRNGYCKXQINGFCV----- 702
QY 91 KQCCPGFSESGEMC---VPHC-ADKCYHRC---IAPWCCCEPGMGSTNC---SSACD 140
DB 703 ---CPDGYHD-HMCLSEVNECSNPCIHGACHDGVNGYKCCCEMGSGSNCNDINNNECE 757
QY 141 GDHMGPHCTSRCCRNKALCNPIITGA--CHCAAGFRGNCEDRCBOGTYYGNDCHORCOCO 198

```

```

Db 758 SN-----PCMGCTCKMTGAVICTKAGFSGPNQ-----TINECSSN-PCL 800
QY 199 NGATG-DHTYG-BCRCPGTYGAFCEDLCPGRKHGPCQECORC---PCONGVCHH-----V 249
Db 801 NHGTCIDVAGYKNCNMLPRTGALICEVLA-----CAGSPCKNGRCKESDPE 850
QY 250 TEGCSCPSGMYGCGPCPEGRFKNCSEC---QCHNGCTDAATG---QCHCSPGYTG 304
Db 851 TFCGCEPFGMGQGTG-----EIDMNECVNRPDRNATCQONTNGSKCKCKPGYTG 900
QY 305 ERQO---DECPVGTGYLCAETCCQVNGKCYHNSGA---CLCEPAGFAGECEARL----- 354
Db 901 RNEEMIDDC-----QPNCHNGSGSCSGINMFNCNCFRAGFRGPCEDINECAS 950
QY 355 -----CPEGLYGIKCDKRCPCHEL---NTHSCHPMG---ECAC 387
Db 951 NPCKNAGNCTCVNSYTCQPGFSGIHCHESNTPDCTESSCFNGTNC---IDGINTFTQCQ 1008
QY 388 KPGWSGLYC---NE-----TCSGFTYGEACQOI---CS---CQ 416
Db 1009 PPEFTSYCOHDINECDSKPCLNGGTCQDSYGTGKCTCPQGYTGLNQNLYRNCDSPPCK 1068
QY 417 NGADCSVTG---KCTCAPGFKGIDCSNP-----442
Db 1069 NGKRCMOTNNFYRCECKSGMTGYCDVPYSCEVAKQGVYDVLICRNSGMCVDTGNTH 1128
QY 443 ---CPLATYGINCSR---CG---CKNDVACSPEVD---SCTCRAGMHVDCS-----483
Db 1129 FCRCQAGYTGSCBEQVDECSFNPQNGATCTDYLGGSECVAGYHVCSEINCLSL 1188
QY 484 -----IRPSTGWGFCNLT---C-----QCLNGACNLTLDG 512
Db 1189 HPCQNGCTCIDLINTYKCSCPRGTOGHCEINVDCTPFYDSFTLBPCKFNKCKIDRVG 1248
QY 513 ---TCTCAPGMRGEKE-----LPCQD---GTYGLNCAE-----RCDC-----SH 546
Db 1249 GYNCICPPGVGERCEGDVNECLSNPCDSRGTG---NCIQLVNDYRCECRGFTGRCESEV 1306
QY 547 ADGC-----HPTTGH-CRCLPGMSGVHCD-----569
Db 1307 VDCCCKMPCRNCGTCANVASNTERGFICKCPGPDGATCEYDSRCSNLCRNGGTCTISVL 1366
QY 570 ----SYCABGRMGPNCSLP 584
Db 1367 TSSKVCSEBGTGATCQYP 1385

```

RESULT 3

```

NTCL_BRAVE
AC NTCL_BRAVE STANDARD: PRT: 2437 AA.
P46350:
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Neurogenic locus notch homolog protein 1 precursor.
GN NOTCH1 OR NOTCH.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=94128602; PubMed=8297791;
RA Bierkamp C., Campos-Ortega J.A.;
RT "A zebrafish homologue of the Drosophila neurogenic gene Notch and
its pattern of transcription during early embryogenesis.";
RL Mech. Dev. 43:87-100(1993).
CC -1- FUNCTION: IMPLICATED IN CELL FATE SPECIFICATIONS DURING
CC EMBRYO DEVELOPMENT. MAY BE INVOLVED IN THE FORMATION OF THE
CC NEURAL PLATE, NOTOCHORD AND BRAIN VESICLES.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.

```

```

CC -1- DEVELOPMENTAL STAGE: Expressed in all cells in pregastrulation
CC stages. During gastrulation is differentially expressed,
CC accumulating predominantly in the prechordal mesoderm and
CC notochord. At the end of gastrulation, expressed along the
CC anterior-posterior axis including the developing neural plate
CC and differentiating mesoderm. Also present in the developing
CC brain and head regions.
CC -1- SIMILARITY: BELONGS TO THE NOTCH FAMILY.
CC -1- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC -1- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.slb.ch).
CC
CC EMBL: X69088; CAA48831.1; -.
CC HSSP: P00740; 1EDM.
CC
CC DR ZEIN: ZDB-GENE-990415-173; notchla.
CC
CC DR InterPro: IPR002110; ANK.
CC DR InterPro: IPR000152; Asx_hydroxyl.
CC DR InterPro: IPR000561; EGF-like.
CC DR InterPro: IPR000742; EGF-2.
CC DR InterPro: IPR001881; EGF_CA.
CC DR InterPro: IPR001438; EGF-II.
CC DR InterPro: IPR000800; Notch.
CC DR Pfam: PF00008; EGF_36.
CC DR Pfam: PF00023; ank; 6.
CC DR Pfam: PF00066; notch; 3.
CC DR PRINTS: PR00010; EGFBL00D.
CC DR PRINTS: PR01452; NOTCH.
CC DR SMART: SM00248; ANK. 5.
CC DR SMART: SM00179; EGF_CA; 19.
CC DR SMART: SM00001; EGF_Like; 16.
CC DR SMART: SM00004; NL; 3.
CC DR PROSITE: PS50088; ANK_REPEAT; 4.
CC DR PROSITE: PS50297; ANK_REPEAT_REGION; 1.
CC DR PROSITE: PS00010; ASX_HYDROXYL; 23.
CC DR PROSITE: PS00022; EGF_1; 34.
CC DR PROSITE: PS01186; EGF_2; 28.
CC DR PROSITE: PS01187; EGF_CA; 22.
CC DR Receptor: Transcription regulation; Activator; Differentiation;
CC KW developmental protein; Neurogenesis; Repeat; ANK repeat;
CC EGF-like domain; Transmembrane; Glycoprotein; Signal.
CC
CC FT SIGNAL 1 20
CC FT CHAIN 21 2437
CC FT DOMAIN 21 1724
CC FT TRANSMEM 1725 1747
CC FT DOMAIN 1748 2437
CC FT DOMAIN 21 57
CC FT DOMAIN 58 98
CC FT DOMAIN 101 138
CC FT DOMAIN 139 175
CC FT DOMAIN 177 215
CC FT DOMAIN 217 254
CC FT DOMAIN 256 292
CC FT DOMAIN 294 332
CC FT DOMAIN 334 370
CC FT DOMAIN 371 409
CC FT DOMAIN 411 449
CC FT DOMAIN 451 487
CC FT DOMAIN 489 524
CC FT DOMAIN 526 562
CC FT DOMAIN 564 599
CC FT DOMAIN 601 637
CC FT DOMAIN 639 674
CC FT DOMAIN 676 712
CC FT DOMAIN 714 749
CC FT DOMAIN 751 787

```

```
FT DOMAIN 789 825 EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 865 865 EGF-LIKE 22, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 867 903 EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 905 941 EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 943 979 EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 981 1017 EGF-LIKE 26, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1019 1055 EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1057 1093 EGF-LIKE 28, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1095 1141 EGF-LIKE 29, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1143 1179 EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1181 1217 EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1219 1263 EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1265 1303 EGF-LIKE 33, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1305 1344 EGF-LIKE 34, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1346 1382 EGF-LIKE 35, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1385 1423 EGF-LIKE 36, CALCIUM-BINDING (POTENTIAL).
FT REPEAT 1446 1486 LIN/NOTCH 1.
FT REPEAT 1487 1520 LIN/NOTCH 2.
FT REPEAT 1521 1561 LIN/NOTCH 3.
FT REPEAT 1567 1910 ANK 1.
FT REPEAT 1915 1944 ANK 2.
FT REPEAT 1948 1978 ANK 3.
FT REPEAT 1982 2011 ANK 4.
FT REPEAT 2015 2044 ANK 5.
FT REPEAT 2048 2077 ANK 6.
FT DOMAIN 2276 2276 POLY-GLN (OPA-REPEAT).
FT SITE 1656 1657 CLEAVAGE BY FURIN-LIKE PROTEASE (BY SIMILARITY).
FT DISULFID 25 35 BY SIMILARITY.
FT DISULFID 29 45 BY SIMILARITY.
FT DISULFID 47 56 BY SIMILARITY.
FT DISULFID 62 73 BY SIMILARITY.
FT DISULFID 67 86 BY SIMILARITY.
FT DISULFID 88 97 BY SIMILARITY.
FT DISULFID 105 116 BY SIMILARITY.
FT DISULFID 110 126 BY SIMILARITY.
FT DISULFID 128 137 BY SIMILARITY.
FT DISULFID 143 154 BY SIMILARITY.
FT DISULFID 148 163 BY SIMILARITY.
FT DISULFID 165 174 BY SIMILARITY.
FT DISULFID 181 194 BY SIMILARITY.
FT DISULFID 188 203 BY SIMILARITY.
FT DISULFID 205 214 BY SIMILARITY.
FT DISULFID 221 232 BY SIMILARITY.
FT DISULFID 242 253 BY SIMILARITY.
FT DISULFID 260 271 BY SIMILARITY.
FT DISULFID 265 280 BY SIMILARITY.
FT DISULFID 282 291 BY SIMILARITY.
FT DISULFID 311 320 BY SIMILARITY.
FT DISULFID 322 331 BY SIMILARITY.
FT DISULFID 338 349 BY SIMILARITY.
FT DISULFID 343 358 BY SIMILARITY.
FT DISULFID 360 369 BY SIMILARITY.
FT DISULFID 375 386 BY SIMILARITY.
FT DISULFID 397 397 BY SIMILARITY.
FT DISULFID 399 408 BY SIMILARITY.
FT DISULFID 415 428 BY SIMILARITY.
FT DISULFID 437 448 BY SIMILARITY.
FT DISULFID 448 466 BY SIMILARITY.
FT DISULFID 460 475 BY SIMILARITY.
FT DISULFID 477 486 BY SIMILARITY.
FT DISULFID 493 503 BY SIMILARITY.
FT DISULFID 498 512 BY SIMILARITY.
FT DISULFID 514 523 BY SIMILARITY.
FT DISULFID 541 550 BY SIMILARITY.
FT DISULFID 553 561 BY SIMILARITY.
FT DISULFID 568 578 BY SIMILARITY.
FT DISULFID 573 587 BY SIMILARITY.
FT DISULFID 589 598 BY SIMILARITY.
```

```
FT DISULFID 605 616 BY SIMILARITY.
FT DISULFID 610 625 BY SIMILARITY.
FT DISULFID 627 636 BY SIMILARITY.
FT DISULFID 643 653 BY SIMILARITY.
FT DISULFID 648 662 BY SIMILARITY.
FT DISULFID 664 673 BY SIMILARITY.
FT DISULFID 680 691 BY SIMILARITY.
FT DISULFID 685 700 BY SIMILARITY.
FT DISULFID 702 711 BY SIMILARITY.
FT DISULFID 718 728 BY SIMILARITY.
FT DISULFID 723 737 BY SIMILARITY.
FT DISULFID 739 748 BY SIMILARITY.
FT DISULFID 755 766 BY SIMILARITY.
FT DISULFID 760 775 BY SIMILARITY.
FT DISULFID 777 786 BY SIMILARITY.
FT DISULFID 793 804 BY SIMILARITY.
FT DISULFID 798 813 BY SIMILARITY.
FT DISULFID 815 824 BY SIMILARITY.
FT DISULFID 831 842 BY SIMILARITY.
FT DISULFID 836 853 BY SIMILARITY.
FT DISULFID 855 864 BY SIMILARITY.
FT DISULFID 871 882 BY SIMILARITY.
FT DISULFID 876 891 BY SIMILARITY.
FT DISULFID 893 902 BY SIMILARITY.
FT DISULFID 909 920 BY SIMILARITY.
FT DISULFID 914 929 BY SIMILARITY.
FT DISULFID 931 940 BY SIMILARITY.
FT DISULFID 947 958 BY SIMILARITY.
FT DISULFID 952 967 BY SIMILARITY.
FT DISULFID 969 978 BY SIMILARITY.
```

Query Match 19.9% Score 717; DB 1; Length 2437;

Best Local Similarity 25.7%; Pred. No. 1.7e-38; Matches 221; Conservative 60; Mismatches 208; Indels 370; Gaps 49;

```
QY 4 SLNSCLSS-----FICLLCHMGTASPLNLEDPNVCSHMESYSYTVORSY 48
DB 601 NINECLSQCRNGRGCGDRENAVICTCPKGTGVNCEINIID--CKR----- 645
QY 49 PHPEQIYITSTDLNNEFKTRHRYSTRATVRHBEKTYRRKSQCCPEFTYSGEMC--- 105
DB 646 -KPCD--YKCIDIDINGEYV-----CEPEY--SGSMCNIN 676
QY 106 VPHCA-----DKCVHRC---IAPWT 123
DB 677 IDDCALNCHNGTGTIDVNSTCTCPDGFPRATCLSOHNESSNPCTHSGSLDINSR 736
QY 124 COCEPFGWGTNCSACDGDHMGPHCTSRCCNGALCNPTGA--CHCAAGFGMRCEDR 181
DB 737 CYCEAGWGMGRNDININ-----ECLSN-PCVNGGCKMDTSGYLCTGRAGFSGPNCOWN 789
QY 182 CEQGTGYGNDCHORCCONGATC-DHYTG-ECCKPFGYTGAFCEDELCPKKGHPQCEORCP 239
DB 790 I-----NECASN-PCLNQSGCIDVDVAGFKCKMPLYGEVCEENVLAP-----CSPR-P 835
QY 240 CONGVYCHN-----VTGEGSCPSGMMGTVCQPCPGRFRFGKNCOSRC---QCHNGGTCDNA 292
DB 836 CKNGGVCRSESEFQSFSCNCPAGMOGQCEVNI-----NECYRRNPGCINGVECENTL 885
QY 293 TG--OCHSPGYTGERQ--DECPVGTGYVLAFTCOQVNGKCY-HVSG-ACICEAGF 345
DB 886 RCGFQCRCPNPGTGTALCENDIDDC-----BPNPCNSGVQCDRVNGFVVCVLAGEF 935
QY 346 AGEBCPEARL-----CPBGLYGTGCDKRCRCHLENTSCHP- 380
DB 936 RGERCAEDIDECVSAPCRNAGNCTDCVNSTGSCPAFGSGINCEINTPCTES--SCENG 993
QY 381 -----MSGECACRPGMSGLYC---NE-----TCSPEFYGEA 408
DB 994 GTCVNGIISFSCVCLPGLFGNTCQHDVNECDSPPCONGSCODGYGTXYKCTCPHGYTGLN 1053
QY 409 COOI---CS---CONGADC--DSVTGKCTCAPGFGIDCSTP----- 442
```

Db 1054 COSLVRWCSSPECKNGSGSCMOQASFTCCQASGNTGICDVPVSCEVAARQGVAVL 1113
 QY 443 -----CPGLTYGINSRCRG-----CKNDAYCSPVDG--SCTCKAGW 477
 Db 1114 CRHAGCCVDAGNTHLCKCAQGYGYSYQEDVDEQPNPCNGATCTTYLGYSCECPGY 1173
 QY 478 HGVDGS-----IRCPSTWFGCNL--TC-----499
 Db 1174 HGMNCSKEINETSQPCQNGTCTDVLVNTYKSCSPRGTQVHECIDIDCSPSVPLTGE 1233
 QY 500 -OCLNGAGCNTLDG--TCTCAPGMRGEKE-----LPCQ-DGTGLMCAE-----RC 542
 Db 1234 PCPENGRCVDRVGVCVCPAGVGRCEGDVNECLSDPCDPSGYS--NCVQLINDRC 1291
 QY 543 DCSHA-----DGHPT-----TGH---CRLPLPGSVYHCD-----569
 Db 1292 ECRTGYGRKCEYVENCCKPTCKNGTCAVASTKRGYICKCPGYSGSSCEYDSGSC 1351
 QY 570 -----SYCAEGRMGFNC 581
 Db 1352 SLRCRNGATCVSGHLSPRC 1370

RESULT 4
 NTG3_HUMAN STANDARD; PRT; 2321 AA.
 ID NTG3_HUMAN
 AC O9UM47; O9Y6L8; O9UPL3; O9UEB3;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Neurogenic locus notch homolog protein 3 precursor (Notch 3).
 GN NOTCH3.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-97032728; PubMed-8878478;
 RA Joutel A., Corpechot C., Ducros A., Vahedi K., Chabrier H., Mouton P.,
 RA Alamowitch S., Domenga V., Cecillon M., Marchal E., MacLazek J.,
 RA Vaysiere C., Cruaud C., Cabanis E.-A., Ruchoux M.M., Weissbach J.,
 RA Bach J.-F., Bousser M.-G., Tournier-Lasserre E.,
 RT "Notch3 mutations in CADASIL, a hereditary adult-onset condition
 causing stroke and dementia.";
 RL Nature 383:707-710(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Guel M., Artavanis-Tsakonas S.;
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Lamerdin J.E., McCreedy P.M., Skowronski E., Adamson A.W.,
 RA Burkhardt-Schultz K., Gordon L., Kyle A., Ramirez M., Stillagen S.,
 RA Phan H., Velasco N., Ganes J., Danganan L., Poundstone P.,
 RA Christensen M., Georgescu A., Avila J., Liu S., Attix C., Andreise T.,
 RA Traubheim M., Amico-Keller G., Coefield J., Duarte S., Lucas S.,
 RA Bruce R., Thomas P., Quan G., Krommiller B., Arellano A.,
 RA Montgomery M., Ow D., Nolan M., Trong S., Kobayashi A., Olsen A.S.,
 RA Carraro A.V.;
 RT "Sequence analysis of an 1.5 Mb olfactory receptor (OLFR) cluster in
 19p13.1";
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP VARIANTS CADASIL Y-49; C-71; C-90; C-110; C-133; C-141; R-146; C-153;
 RP C-169; C-171; C-182; R-185; S-212; G-222; Y-224; C-258; Y-542; C-558;
 RP C-578; C-728; C-985; C-1006; C-1031; C-1231 AND R-1261, AND VARIANTS
 RP R-170; L-496; Q-1133; M-1183 AND A-2223.
 RX MEDLINE-98049753; PubMed-9388399;
 RA Joutel A., Vahedi K., Corpechot C., Troesch A., Chabrier H.,
 RA Vaysiere C., Cruaud C., MacLazek J., Weissbach J., Bousser M.-G.,
 RA Bach J.-F., Tournier-Lasserre E.;
 RT "Strong clustering and stereotyped nature of Notch3 mutations in

RT CADASIL patients.";
 RL Lancet 350:1511-1515(1997).
 RN [5]
 RP VARIANT CADASIL 114-GLY--PRO-120 DEL.
 RX MEDLINE-20264473; PubMed-10802807;
 RA Joutel A., Chabrier H., Vahedi K., Domenga V., Vaysiere C.,
 RA Ruchoux M.M., Lucas C., Leys D., Bousser M.-G., Tournier-Lasserre E.;
 RT "Splice site mutation causing a seven amino acid Notch3 in-frame
 deletion in CADASIL.";
 RL Neurology 54:1874-1875(2000).
 RN [6]
 RP IDENTIFICATION OF LIGANDS.
 RX MEDLINE-99180765; PubMed-10079256;
 RA Gray G.E., Mann R.S., Mitsiadis E., Henrique D., Carcangu M.-L.,
 RA Banks A., Leiman J., Ward D., Ish-Horowitz D., Artavanis-Tsakonas S.;
 RT "Human ligands of the Notch receptor.";
 RL Am. J. Pathol. 154:785-794(1999).
 CC -1- FUNCTION: Functions as a receptor for membrane-bound ligands
 CC Jagged1, Jagged2 and Delta to regulate cell-fate determination.
 CC Upon ligand activation through the released notch intracellular
 CC domain (NICD) it forms a transcriptional activator complex with
 CC RBP-J kappa and activates genes of the enhancer of split locus.
 CC Affects the implementation of differentiation, proliferation and
 CC apoptotic programs (by similarity).
 CC -1- SUBUNIT: Heterodimer of a C-terminal fragment N(TW) and a N-
 CC terminal fragment N(EC) which are probably linked by disulfide
 CC bonds (By similarity).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Following
 CC proteolytical processing NICD is translocated to the nucleus.
 CC -1- TISSUE SPECIFICITY: Ubiquitously expressed in fetal and adult
 CC tissues.
 CC -1- PTM: Synthesized in the endoplasmic reticulum as an inactive form
 CC which is proteolytically cleaved by a furin-like convertase in the
 CC trans-Golgi network before it reaches the plasma membrane to yield
 CC an active ligand-accessible form. Cleavage results in a C-
 CC terminal fragment N(TW) and a N-terminal fragment N(EC). Following
 CC ligand binding, it is cleaved by TNF-alpha converting enzyme
 CC (TACE) to yield a membrane-associated intermediate fragment called
 CC notch extracellular truncation (NEXT). This fragment is then
 CC cleaved by presenilin dependent gamma-secretase to release a
 CC notch-derived peptide containing the intracellular domain (NICD)
 CC from the membrane (By similarity).
 CC -1- PTM: Phosphorylated (By similarity).
 CC -1- DISEASE: Defects in NOTCH3 are associated with CADASIL (cerebral
 CC autosomal dominant arteriopathy with subcortical infarcts and
 CC leukoencephalopathy) which causes a type of stroke and dementia of
 CC which key features include recurrent subcortical ischemic events
 CC and vascular dementia.
 CC -1- SIMILARITY: BELONGS TO THE NOTCH FAMILY.
 CC -1- SIMILARITY: CONTAINS 34 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
 CC -1- SIMILARITY: CONTAINS 5 ANK REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; U97669; AAB91371.1; -
 DR EMBL; AF058900; AAC14346.1; -
 DR EMBL; AF058881; AAC14346.1; JOINED.
 DR EMBL; AF058882; AAC14346.1; JOINED.
 DR EMBL; AF058883; AAC14346.1; JOINED.
 DR EMBL; AF058884; AAC14346.1; JOINED.
 DR EMBL; AF058885; AAC14346.1; JOINED.
 DR EMBL; AF058886; AAC14346.1; JOINED.
 DR EMBL; AF058887; AAC14346.1; JOINED.
 DR EMBL; AF058888; AAC14346.1; JOINED.
 DR EMBL; AF058889; AAC14346.1; JOINED.
 DR EMBL; AF058890; AAC14346.1; JOINED.

DR EMBL: AF058891; AAC14346.1; JOINED.
 DR EMBL: AF058892; AAC14346.1; JOINED.
 DR EMBL: AF058893; AAC14346.1; JOINED.
 DR EMBL: AF058894; AAC14346.1; JOINED.
 DR EMBL: AF058895; AAC14346.1; JOINED.
 DR EMBL: AF058896; AAC14346.1; JOINED.
 DR EMBL: AF058897; AAC14346.1; JOINED.
 DR EMBL: AF058898; AAC14346.1; JOINED.
 DR EMBL: AF058899; AAC14346.1; JOINED.
 DR EMBL: AC004257; AAC04897.1; JOINED.
 DR EMBL: AC004663; AAC15789.1; ALT_INIT.
 DR HSSP: P00740; 1EDM.
 DR Genew: HGNC:7883; NOTCH3.
 DR MIM: 125310; -.
 DR InterPro: IPR002110; ANK.
 DR InterPro: IPR000152; ASX_hydroxyl.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR000742; EGF-2.
 DR InterPro: IPR001881; EGF_Ca.
 DR InterPro: IPR001438; EGF_II.
 DR InterPro: IPR002049; Laminin_EGF.
 DR InterPro: IPR008000; Notch.
 DR Pfam: PF00008; EGF; 34.
 DR Pfam: PF00023; ank; 6.
 DR Pfam: PF00066; notch; 3.
 DR PRINTS: PR00010; EGFBLDOD.
 DR PRINTS: PR00011; EGFAMININ.
 DR PRINTS: PR01452; NOTCH.
 DR SMART: SM00248; ANK; 5.
 DR SMART: SM00179; EGF_CA; 19.
 DR SMART: SM00001; EGF_Like; 14.
 DR SMART: SM00004; NL; 3.
 DR PROSITE: PS50088; ANK_REPEAT; 4.
 DR PROSITE: PS50297; ANK_REPEAT_REGION; 1.
 DR PROSITE: PS00010; ASX_HYDROXYL; 18.
 DR PROSITE: PS00022; EGF_1; 33.
 DR PROSITE: PS01186; EGF_2; 25.
 DR PROSITE: PS01187; EGF_CA; 18.
 KW Receptor: Transcription regulation; Activator; Differentiation;
 KW Developmental protein; Repeat; ANK repeat; EGF-like domain;
 KW Transmembrane; Glycoprotein; Signal; Phosphorylation; Polymorphism;
 KW Disease mutation.
 FT SIGNAL 1 39 POTENTIAL.
 FT CHAIN 40 2321 NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 3.
 FT CHAIN 1629 2321 NOTCH EXTRACELLULAR TRUNCATION (BY
 FT CHAIN 1662 2321 SIMILARITY).
 FT CHAIN 1643 1643 NOTCH INTRACELLULAR DOMAIN (BY
 FT TRANSMEM 1644 1664 EXTRACELLULAR (POTENTIAL).
 FT 1665 2321 POTENTIAL.
 FT 40 77 CYTOPLASMIC (POTENTIAL).
 FT 78 118 EGF-Like 1.
 FT 119 156 EGF-Like 2.
 FT 158 195 EGF-Like 3.
 FT 197 234 EGF-Like 4.
 FT 236 272 EGF-Like 5.
 FT 274 312 EGF-Like 6.
 FT 314 350 EGF-Like 7.
 FT 351 389 EGF-Like 8.
 FT 391 429 EGF-Like 9.
 FT 467 495 EGF-Like 10.
 FT 507 543 EGF-Like 11.
 FT 545 580 EGF-Like 12.
 FT 582 618 EGF-Like 13.
 FT 620 655 EGF-Like 14.
 FT 657 693 EGF-Like 15.
 FT 695 730 EGF-Like 16.
 FT 734 770 EGF-Like 17.
 FT 771 808 EGF-Like 18.
 FT 810 847 EGF-Like 19.
 DOMAIN 810 847 EGF-Like 20.
 DOMAIN 810 847 EGF-Like 21.
 DOMAIN 810 847 CALCIUM-BINDING (POTENTIAL).

FT DOMAIN 849 885 EGF-Like 22, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 887 922 EGF-Like 23, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 924 960 EGF-Like 24.
 FT DOMAIN 962 998 EGF-Like 25.
 FT DOMAIN 1000 1034 EGF-Like 26.
 FT DOMAIN 1047 1082 EGF-Like 27.
 Query Match 19.4%; Score 697; DB 1; Length 2321;
 Best Local Similarity 25.2%; Pred. No. 3e-37;
 Matches 226; Conservative 61; Mismatches 249; Indels 360; Gaps 51;
 QY 5 INSCLS-----FICLLCHMIGTASPLNED-----PNVC-SHWESY 40
 DB 432 VNECLSGPCRNATCLDRIGFTCTCMAGFTGYEVDIDEGQSPCVNGCKRNVNGF 491
 QY 41 SYTVQESTPHPPDQIYYTSC--TDILNFKCTRNAVSTRATVHGEKTMTRKSCCCGF 98
 DB 492 SCTCPSGFSGSTCQLDVDECASTPCRNAGKACVDPDGY-----ECRCAGGF 537
 QY 99 YESGEMC---VPHCA-DKCVHGRC---IAPNTCCQEPGMGNCSSACDGHMGPCTSR 151
 DB 538 --EGTLCRRNDDCSPDPCHHRCYDGIASFSACAPGTGTCSQYD-----ECRSQ 589
 QY 152 CQCKNGALCNPTTG--ACHCAAGFRGWCE--DRCEQG--TYG--NDCHOR--CQCGNG 200
 DB 590 -PCHRGKCLDLYDKYLCRPSGTGVNCEVNIIDCASNPTCFGVCRDGINRYDCVCPG 648
 QY 201 AT---CDHYTGRCRCPPTGTAFCD-----LCPGKHGQC---EQRC---PCQNG 244
 DB 649 FTGPLCNVEINECASSPCGEGSCVDGNGFRCLCPGLPLCLPSPHPCAHPCSH-G 707
 QY 245 VCHHTG--ECSGPGMMGTVCQ-----PCPEGRGK 275
 DB 708 ICYDAPGFRVCYCEPBGNGSPRCSQSLADACSPCRAGTSSDGMGFHCTCPGVGR 767
 QY 276 NCS--QEC--QCHNGTCDATGQ--CHSPGYTGRGQ--DEC-----PVGTGYVLC 320
 DB 768 OCELLSPCTPNPNCHEGRCEASAPGLPYCSCPGQNGPRCQDDVDECAAGAPACGPHGT-C 826
 QY 321 AE-----TCQ-----CVNGKCYHNSG--ACICEAGFAEREA 352
 DB 827 TNLGFSFCTCHGTYGTPSCDDINDCDPNPLNGSCODGVGSPSCCLPFAFPRC-A 885
 QY 353 R-----LCEPGLYIGICDKRCPCHLNTSHCHM----- 381
 DB 886 RDVDECLNPGCPGCTGTTHVASFCTCPRGYGFHCEDDL-----DCSPSSCFNGCT 938
 QY 382 -----SGEACAKPMSGLYCNE-----TCSPGFYGACQ 410
 DB 939 CVDGVNSFSCLCRPGYTGAGHQHEADPCLSRCLHGVCSAHPGFRCCTCLESFTGPCCQ 998
 QY 411 QI---CS---CONGADCSVTGKCTCAPGFKIDC---STP----- 442
 DB 999 TLYDMCSROPQONGRCYVOTGAYCLCPGWSGRLDLSLPCREANAAGVLEQLCOAG 1058
 QY 443 -----CPLTGYGNCSSRCG-----CKNAVCSPVDS--CTCKAGHWGD 481
 DB 1059 GOCVDEDSHYCYCPGRTGSHCEQVDPCLAQPCQHGCTGTYGTCBCLPGLNDN 1118
 QY 482 CS-----IRCSGTGFCNLT--C-----QCLN 503
 DB 1119 CEDVDDECAAPQCGGSCIDLVARLYLSCPGTGLGYCEINEDCGGPPPLDSGPRCLH 1178
 QY 504 GGACNTLDG--TCTCAPRWGKEKEL---PCQDGTYGLANCERCCSHADGCHPTTG--- 555
 DB 1179 NGTCVDLGVGFRCTCPGTYTGLRCADINEGRSSG-----CHAATTRCLDPPGGGF 1230
 QY 556 HCRCLPGMSGVHCDV-----CAEGRMGPRC 581
 DB 1231 RCLCHAGSFGPRCQTVLSPCESQPCQHGQCRPSPGPGGLFTYCHCAQPMWPRC 1286
 RESULT 5

NC1_MOUSE STANDARD; PRT: 2531 AA.
 ID NC1_MOUSE
 AC 001705; 09QW58; 09QJ23; 006007; 061905; 09R0X7;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Neuregulin locus notch homolog protein 1 precursor (Notch 1) (Molch A) (mrl4) (p300).
 OS NOTCH1 OR NOTCH.
 GN Mus musculus (Mouse).
 OC Eukaryota; Euteleostomi; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC TISSUE-Embryo;
 RX MEDLINE-93194170; PubMed-8449489;
 RA Franco del Amo F., Gendron-Maguire M., Swiatek P.J., Jenkins N.A., Copeland N.G., Gridley T.;
 RT "Cloning, analysis, and chromosomal localization of Notch-1, a mouse homolog of Drosophila Notch.";
 RL Genomics 15:259-264(1993).
 RN [2]
 RP SEQUENCE OF 731-1899 FROM N.A. (ISOFORM 2), AND DEVELOPMENTAL STAGE.
 RC STRAIN-CD-1; TISSUE-Embryo;
 RX MEDLINE-93050801; PubMed-1426644;
 RA Reaume A.G., Conlon R.A., Zingibbi R., Yamaguchi T.P., Rossant J.;
 RT "Expression analysis of a Notch homologue in the mouse embryo.";
 RL Dev. Biol. 154:377-387(1992).
 RN [3]
 RP SEQUENCE OF 1551-1647 FROM N.A. (ISOFORM 1), AND DEVELOPMENTAL STAGE.
 RC TISSUE-Embryo;
 RX MEDLINE-9304835; PubMed-1425352;
 RA Franco del Amo F., Smith D.E., Swiatek P.J., Gendron-Maguire M., Greenspan R.J., McMahon A.P., Gridley T.;
 RT "Expression pattern of Notch, a mouse homolog of Drosophila Notch, suggests an important role in early postimplantation mouse development.";
 RL Development 115:737-744(1992).
 RN [4]
 RP SEQUENCE OF 1161-1547 FROM N.A.
 RC STRAIN-C57BL/6 X CBA; TISSUE-Embryo;
 RX MEDLINE-93178563; PubMed-8440332;
 RA Lardelli M., Lendahl U.;
 RT "Morch A and Notch B-two mouse Notch homologues coexpressed in a wide variety of tissues.";
 RL Exp. Cell Res. 204:364-372(1993).
 RN [5]
 RP SEQUENCE OF 1659-1673 FROM N.A.
 RX MEDLINE-99364499; PubMed-10437788;
 RA Lee J.S., Ishimoto A., Yanagawa S.I.;
 RT "Murine leukemia provirus-mediated activation of the Notch1 gene leads to induction of HES-1 in a mouse T lymphoma cell line, DL-3.";
 RL FEBS Lett. 455:276-280(1999).
 RN [6]
 RP SEQUENCE OF 1950-2201 FROM N.A.
 RX MEDLINE-98029496; PubMed-9384671;
 RA Messierie M., Folio M., Nehls M., Eggert H., Boehm T.;
 RT "Dynamic changes in gene expression during in vitro differentiation of mouse embryonic stem cells.";
 RL Cytokines Cell. Mol. Ther. 1:139-143(1995).
 RN [7]
 RP SEQUENCE OF 1555-1659, CLEAVAGE BY FURIN-LIKE CONVERTASE, AND MUTAGENESIS OF 1651-ARG-GLN-ARG-ARG-1654.
 RX MEDLINE-98318619; PubMed-9653148;
 RA Logeat F., Bessia C., Brou C., Leball O., Jarriault S., Seidah N.G., Israel A.;
 RT "The Notch1 receptor is cleaved constitutively by a furin-like convertase.";
 RL Proc. Natl. Acad. Sci. U.S.A. 95:8108-8112(1998).
 RN [8]
 RP PARTIAL SEQUENCE, AND POST-TRANSLATIONAL PROCESSING.
 RX MEDLINE-21523956; PubMed-11518718;

RA Saxena M.T., Schroeter E.H., Mumm J.S., Kopan R.;
 RT "Murine notch homologs (NL-4) undergo presenilin-dependent proteolysis.";
 RL J. Biol. Chem. 276:40268-40273(2001).
 RN [9]
 RP POST-TRANSLATIONAL PROCESSING.
 RX MEDLINE-21374376; PubMed-11459941;
 RA Mizutani T., Taniguchi Y., Aoki T., Hashimoto N., Honjo T.;
 RT "Conservation of the biochemical mechanisms of signal transduction among mammalian Notch family members.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9026-9031(2001).
 CC -1- FUNCTION: Functions as a receptor for membrane-bound ligands Jagged1, Jagged2 and Delta to regulate cell-fate determination. upon ligand activation through the released notch intracellular domain (NICD) it forms a transcriptional activator complex with RBP-1 kappa and activates genes of the enhancer of split locus. affects the implementation of differentiation, proliferation and apoptotic programs (by similarity). May play an essential role in postimplantation development, probably in some aspect of cell specification and/or differentiation. May be involved in mesoderm development, somite formation and neurogenesis. Involved in the maturation of both CD4+ and CD8+ cells in the thymus.
 CC -1- SUBUNIT: Heterodimer of a C-terminal fragment N(Tw) and a N-terminal fragment N(Ec) which are probably linked by disulfide bonds.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Following proteolytical processing NICD is translocated to the nucleus.
 CC -1- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; may be produced by alternative splicing.
 CC -1- TISSUE SPECIFICITY: Highly expressed in the brain, lung and thymus. Expressed at lower levels in the spleen, bone-marrow, spinal cord, eyes, mammary gland, liver, intestine, skeletal muscle, kidney and heart.
 CC -1- DEVELOPMENTAL STAGE: First detected in the mesoderm at 7.5 d.p.c. By 8.5 d.p.c. highly expressed in presomitic mesoderm, mesenchyme and endothelial cells, while much lower levels are seen in the neuroepithelium. Between 9.5-10.5 d.p.c. expressed at high levels in the neuroepithelium. At 13.5 d.p.c. expressed in the surface ectoderm, eye and developing whisker follicles.
 CC -1- PTM: Synthesized in the endoplasmic reticulum as an inactive form which is proteolytically cleaved by a furin-like convertase to yield trans-golgi network before it reaches the plasma membrane to yield an active, ligand-accessible form. Cleavage results in a C-terminal fragment N(Tw) and a N-terminal fragment N(Ec). Following ligand binding, it is cleaved by TNF-alpha converting enzyme (TACE) to yield a membrane-associated intermediate fragment called notch extracellular truncation (NEXT). This fragment is then cleaved by presenilin dependent gamma-secretase to release a notch-derived peptide containing the intracellular domain (NICD) from the membrane.
 CC -1- PTM: Phosphorylated.
 CC -1- SIMILARITY: BELONGS TO THE NOTCH FAMILY.
 CC -1- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 3 LIN/NORCH REPEATS.
 CC -1- SIMILARITY: CONTAINS 5 ANK REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: 211886; CAA7941.1; -
 CC EMBL: 102613; AAK14898.1; -
 CC EMBL: X68278; CAA48339.1; -
 CC EMBL: AJ238029; CAB40733.1; -
 CC EMBL: X82562; CAA57909.1; -
 CC HSSP: P00740; 1EDM.
 CC MGD: MGI:97363; Notch1.
 CC InterPro: IPR002110; ANK.
 CC InterPro: IPR000152; Asx_hydroxyl.

Db	760	N-----	PCVNGGTCTCDMTSGVYCTCREFSPPNQ-----	TINNECASN-PCLN	802
Qy	200	GATC-DHVTG-ECCNCPERTYGAFCEDLCPRGKHGPCOEORCPONGVC----	HHVTGEC	253	
Db	803	QGTCTDDVAGYKCNCPPLPYTGATCEVYLA-----	C-ATSPCKNSGVCKESEDYESPSC	855	
Qy	254	SCBPGMNGVYC-----	GQPCPEG-----	FGKNS---QECQ---	282
Db	856	VCPPRMOGQCEVDINECVKSPCRHGASCONTNGSYRCLQAGYTGRCNESDIDCRRNP	915		
Qy	283	CHNGGCTDAA--TGQCHCSGTYGRCODE-----	CPYGT	315	
Db	916	CHNGGCTDGINTAFCDCILPFGAGFCBEDINECASNPPONGANCITDCYDSTTCTCPVGF	975		
Qy	316	YGVYCAET-----	CQCVNGSKCYHYSG-----	ACLCAEAGFASERCEARLCPGLYGI-KC	364
Db	976	NGICEENPPDCTRESSCFNGGTC--VDGINSFCLCPGPTGYSQ-----	YDVNC	1022	
Qy	365	DKRCPCHEINTHSHCFMSSG--EACAKPQMSGLYCNE-----	TCSRQFYEAACQI-----	412	
Db	1026	DSR-PCLLHGT--CODSYGYTKCTCPQGYGLNCONLVKMCDSAPCKNGRCMCTNTQYH	1082		
Qy	413	CSQCN-----	GADCDQSYNGKCTCAGPFGIDCSFPCPLGTGIN-----	CSSRCG-----	457
Db	1083	CECRSGMTGVNCDVLSVSCVLAOKKRGIDTLLQHGGLCVDBGDKHYCHCOAGYTGSTC	1142		
Qy	458	-----	CKNDAYCSFVDS--SCCTKAGHNGYDS-----	483	
Db	1143	EDVEDESRPQCNQAGATCTIYLGFSCKCAVGHGSCSEINECLSPQCPQNGGTCIDLT	1207		
Qy	484	----	IRCPSGTGWFGCNLT--C-----	QCLNGACNTLDG--TTCAPAGMGE	523
Db	1203	NSYKCSPPRGTOGVCHEINVDCHPLDPAISRSPKCFNNGTQVDQYGGYTCPCPGFGE	1262		
Qy	524	KCE-----	LEQOD-GTYGLNCAER--CDC-----	SHAGC-----	550
Db	1263	RCBEDVNECLSNPDCDPRGTD--NCQVRVNDFHCECRAGHTGRRCSEVINGCGKPCCKNGG	1320		
Qy	551	-----	HPITGH--CRCLPQMSGVHCS-----	VCABGRMGPNC-----	581
Db	1321	VCANASTATAGFICRCRCPAGRBGATCENDARTCGSLCLNGNGTCISQPSRPTCLCLGSPFG	1380		
Qy	582	-----	SLPCY	586	
Db	1381	PECFPASSPCVGSNPCY	1398		
RESULT 6					
NTC3_MOUSE					
AC	061982;	STANDARD;	PRT;	2318	AA.
DT	01-NOV-1997 (Rel. 35, Created)				
DT	01-NOV-1997 (Rel. 35, Last sequence update)				
DT	15-JUN-2002 (Rel. 41, Last annotation update)				
DE	Neurogenic locus notch homology protein 3 precursor (Notch 3).				
GN	NOTCH3.				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
OX	NCBI_TaxID:10090;				
RN	SEQUENCE FROM N.A.				
RP	STRAIN-ICR X Swiss Webster;				
RC	MEDLINE-95001556; PubMed-7918097;				
RA	Lardelli M., Dalstrand J., Lendahl U.;				
RT	"The novel Notch homologue mouse Notch 3 lacks specific epidermal				
RT	growth factor-repeats and is expressed in proliferating				
RT	neuroepithelium";				
RL	Mech. Dev. 46:123-136(1994).				
RP	POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS OF MET-1664.				
RX	MEDLINE-21523956; PubMed-11518718;				

FT DISULFID 241 252 BY SIMILARITY.
 FT DISULFID 246 261 BY SIMILARITY.
 FT DISULFID 263 272 BY SIMILARITY.
 FT DISULFID 279 292 BY SIMILARITY.
 FT DISULFID 286 301 BY SIMILARITY.
 FT DISULFID 303 312 BY SIMILARITY.
 FT DISULFID 319 330 BY SIMILARITY.
 FT DISULFID 324 339 BY SIMILARITY.
 FT DISULFID 341 350 BY SIMILARITY.
 FT DISULFID 356 367 BY SIMILARITY.
 FT DISULFID 361 378 BY SIMILARITY.
 FT DISULFID 380 389 BY SIMILARITY.
 FT DISULFID 396 409 BY SIMILARITY.
 FT DISULFID 403 418 BY SIMILARITY.
 FT DISULFID 420 429 BY SIMILARITY.
 FT DISULFID 436 447 BY SIMILARITY.
 FT DISULFID 441 456 BY SIMILARITY.
 FT DISULFID 458 467 BY SIMILARITY.
 FT DISULFID 474 485 BY SIMILARITY.
 FT DISULFID 479 494 BY SIMILARITY.
 FT DISULFID 496 505 BY SIMILARITY.
 FT DISULFID 512 523 BY SIMILARITY.
 FT DISULFID 517 532 BY SIMILARITY.
 FT DISULFID 534 543 BY SIMILARITY.
 FT DISULFID 550 560 BY SIMILARITY.
 FT DISULFID 555 569 BY SIMILARITY.
 FT DISULFID 571 580 BY SIMILARITY.
 FT DISULFID 587 598 BY SIMILARITY.
 FT DISULFID 592 607 BY SIMILARITY.
 FT DISULFID 609 618 BY SIMILARITY.
 FT DISULFID 625 635 BY SIMILARITY.

Query Match 19.0% Score 685.5; DB 1; Length 2318;

Best Local Similarity 24.4% Pred. No. 1.6e-36; Matches 216; Conservative 59; Mismatches 195; Indels 417; Gaps 48;

QY 94 CCGFYEGEMCPHCDKCYHGRCAIAPNT-----CCCEPQMGSTNCSCADGDHW 144
 DB 225 CLPGE--EGNCEVNA-VDDCGHRLNGTCTVDGVTNYNCCCPREMTQFCTEDVD---- 277
 QY 145 GPHCTSRQ-----CKNALNPITG--ACHCAAG-----FRGMRCE 179
 DB 278 -----ECQLPQNMCHNGCTFNLGSHSCVNGMTGESCSQONIDCATAVCFHGATCH 331
 QY 180 DR-----CEGTYGNDCH--QRC--QCONGATCD--HYTGE--CGRPGTYGAPCED 223
 DB 332 DRVASFYCACPMGKTGLLCHLDACVSNPCHEDAICDINPVSGRAICTCPGFTGGACDQ 391
 QY 224 -----LCPGKHGPOCE----- 235
 DB 392 DVDECSIGANPCENHGRVNTGSLCCGGRGTGPRCTEDVNECLSGPCRNQATCLDRI 451
 QY 236 -----QRCPCQNGVC--HHVTG--ECSCPSGMGTVC----- 264
 DB 452 GQFTICMAGFTGYCEVDIDECOSSPCVNGVCKDRVNGFSCITCPSGSGMQLDYDE 511
 QY 265 -----GPR-----CPEGRGKNCQ--ECQ--CHNGSTCA--ATGQCHC 298
 DB 512 CASTPCCRNGANCVDOPDGETEGRCAABGEBGLTCERNVDCSPDPCHGRCVGLASFSCAC 571
 QY 299 SPGYTGERCODE-----CPVTTYGYLC-----AETCQ 325
 DB 572 APGYTGHCESQVDECRSQPRYGRKGLDYDKYLICRPPITTYGNCVNIIDDCASNCT 631
 QY 326 ---CVNGSKCYHVGACLCERFNGERCEARL-----CPEGLYGIKCDKRCR 369
 DB 632 FGVCGRDGINRYD---CVCQPGFTGPLCNEINECCASPCGEGSCVVGENGPHC--LCP 685
 QY 370 -----CHLENTHS--CHPMSG--ECACKRWSGLYCNE----- 398
 DB 686 PGLSPPLCLPANHPCAHFPCSHGVCHDAPGRCVCEGWSGPRCSOSLADPACESOPCQ 745
 QY 399 -----TCSPGFYEGACOOI--CS---CONGADCDSVYGR---CTCAPGFKG 436

DB 746 AGGTCTSDGIGFRCACAGFGHGCCEVLSPTCTPSLCEHGHCEDDPRLTYCSPFPMQGS 805
 QY 437 -----IDCTPDLGTYGINCSS-----ROGCKNDAY-----CSP----- 466
 DB 806 PRCODVDEGACASPCGPHG--CTNLPGNFRCIGHRYTGFPFCDDIDDCPNLHGGS 864
 QY 467 -VDG-----SCYCKXGMHGVDC-----SINCPSTGNGFGONTL 497
 DB 865 CQDVGFSFSCSLDGFAPRCARDVDECLSSPCGPGCTDTHVASFCTACPPYGGFCEI 924
 QY 498 -----TQCCJNGGACNTLIDG-----TCTCAPGMRGEC----- 525
 DB 925 DLDPQSPSCNNGGIC--VDGVSFSFSLCRGRTGYTHQYADPFSRPLHGGICNPTH 982
 QY 526 ---ELPCODGYTYGLNCAERCD-----CSHADGCHPTTHGRCPLDPMGSHVCD----- 569
 DB 983 PGFECTCREGFTGSCOCNPVDMCSQAPCONGRCVOTGATYCICPGWSGRLCDIQSLPCT 1042
 QY 570 -----SYCAEGRWGPNCSTL---PC 585
 DB 1043 EAAAMGVRLBQLCOEGSKCIDKGRSHYCVCPRGHTSGCHEYDPC 1089

RESULT 7

ID NTC2_RAT STANDARD; PRT: 2471 AA.
 AC O9QW30;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Neurogenic locus notch homolog protein 2 precursor (Notch 2).
 GN NOTCH2.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_Taxid=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain.
 RX MEDLINE=93202015; PubMed=1295745;
 RA Weinmaster G., Roberts V.J., Lemke G.;
 RT "Notch2: a second mammalian Notch gene.";
 RL Development 116:931-941(1992).
 RP [2]
 RP TISSUE SPECIFICITY.
 RX MEDLINE=21331789; PubMed=11438922;
 RA Irvin D.R., Zurcher S.D., Nguyen T., Weinmaster G., Kornblum H.I.;
 RT "Expression patterns of Notch1, Notch2, and Notch3 suggest multiple functional roles for the Notch-Delta signaling system during brain development.";
 RL J. Comp. Neurol. 436:167-181(2001).
 CC -1- FUNCTION: Functions as a receptor for membrane-bound ligands Jagged1, Jagged2 and Delta to regulate cell-fate determination. Upon ligand activation through the released notch intracellular domain (NICD) it forms a transcriptional activator complex with RBP-J kappa and activates genes of the enhancer of split locus. Affects the implementation of differentiation, proliferation and apoptotic programs. May play an essential role in postimplantation development, probably in some aspect of cell specification and/or differentiation. (By similarity).
 CC -1- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-terminal fragment N(EC) which are probably linked by disulfide bonds (By similarity).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Following proteolytical processing NICD is translocated to the nucleus.
 CC -1- TISSUE SPECIFICITY: Highly expressed in the spleen and choroid plexus in the brain. Expressed in postnatal central nervous system (CNS) germinal zones and, in early postnatal life, within numerous cells throughout the CNS. It is more highly localized to ventricular germinal zones. Also found in the heart, liver and kidney.
 CC -1- DEVELOPMENTAL STAGE: Expressed in the brain during E14 and E17.

[illegible]

FT DISULFID 498 509 BY SIMILARITY.
 FT DISULFID 503 518 BY SIMILARITY.
 FT DISULFID 520 529 BY SIMILARITY.
 FT DISULFID 536 547 BY SIMILARITY.
 FT DISULFID 541 556 BY SIMILARITY.
 FT DISULFID 558 567 BY SIMILARITY.
 FT DISULFID 574 584 BY SIMILARITY.

Query Match 19.08; Score 685.5; DB 1; Length 2471;
 Best Local Similarity 24.88; Pred. No. 1.7e-36;
 Matches 218; Conservative 83; Mismatches 240; Indels 339; Gaps 55;

QY 3 ISLNSCLFICL---LLCH-----WIGFASPLNE--DPNVCSHW-----ES 39
 Db 531 IDIDCSSTPLNGAKKCIDHNGYRCOCATGFTGLCEINDNCDPDCHHGOCODGIDS 590
 QY 40 YSVTVOESYPHPF--DQI--YVTS-----CTDLNMFECT-----RHRSY---- 76
 Db 591 YTCICNPYGMALCSDOIDECYSPLNDGRICIDLVNGYQCNCGPGTSGLNCIEINFDCA 650
 QY 77 RTAVRHGF--KTYRRKQCCGPFYSESGMC--VPHCADR----- 112
 Db 651 SNPLHGACVDGINKYSCVCSGPF--TGQRNIDIDECASNPCKRDATCINDVNGFRMC 708
 QY 113 -----CVHGRG--IAPNTCCBPNGGFGNCSACDGDHMGPHCT 149
 Db 709 PEGRHPSYCQVNECLSPCIHNGTCIGLSGYKCLDAGVGINCE--VDKN-----ECL 762
 QY 150 SRCQCKNALCNPTGA--CHCAAGFRGWCE--DRC-----EQTYYGNDCH--QRCQC 197
 Db 763 SN-PQONGGTCNNLVNGYRCYCKKGFKNQVNIDECASNPCLNOGCLDVGSGYCHC 821
 QY 198 Q--NGATCDHYTGRCRPPGTGAFCE-----LCPRGKHGPGCE--QRC---P 239
 Db 822 MLPTYGKQCQYVLAFCSPNCPENAVYCKEAPNFESFTCLCAPGOGQCTYVDVDECVSKP 881
 QY 240 QONGVCCHRYTGE--CSPSGMGTVCQPCPEGRFGRKNSQEC--OCHNGGTC--DAA 292
 Db 882 CMNNGICHTNGQSYWCCECPGPGSGMDCEBDI-----NDCLANPCONGSGVDKVN 931
 QY 293 TGGCHSGYTGGERCODE-----CVGYTYGVC--AETC 324
 Db 932 TFSCLCLPFGVGDCKQDTMNBCLSPCKNNGTCSDYVNSTYCTCPAGHGVHCENNIDEC 991
 QY 325 ---OCVNGKCYHVS---ACICEAGFAGERC-----EARLPEGLYGIKC 364
 Db 992 TESSCFNGGTC--VDGINSFSCLCFVGFTGPCLHDINECSSNPCLNSGTGDLGTYRC 1049
 QY 365 -----DKRC-----PCHLENTSHCHPMSGECACRPGMSGLYCNE----- 398
 Db 1050 TCIPLGTGKNCQTLVNLCSPPCKKNGKGCQAQEKARPRCLCPRGMGAGAYCVLANSCKAAA 1109
 QY 399 -----TCSPGFYGEACQ---ICS--CQNGACDSVYG 426
 Db 1110 LQKGVPEHLCONSGICINAGNTNHCQPLGTGSGYCEQDDECCASNPCCQHGATCSDFIG 1169
 QY 427 --KCTCAKFGFIDCSFTPLGTGINSNRCCNDNAVCSVPDSCCKKMGHVDISI 484
 Db 1170 GYRCVCYGVGVNCE-----YVDECCQNPCCNGGTCTIDLNVHFKCS----- 1212
 QY 485 RCPSTGWFQGC--NL--TC-----OCLNGGAC--NTLDG--TCTCAPGWRGKCE-----L 527
 Db 1213 -CPETRBLLEENIDDAAGAPHCINGGQCYDRIGYSCRLPLGFAGRCEBDINECSLN 1271
 QY 528 PC-QDGTGGLNCAE-----KDCDCHA-----DGCH-----PTTG 555
 Db 1272 PCSSGGS--LDCIQLKNNYQCVCSRSAFGRHCEFFLDVCPQKPCLINGSTCAVANVPDGF 1329
 QY 556 HCRCLPGMSGVHCDVCAEGRM-----GRNCSLP 584
 Db 1330 ICRCPGFGSGARCOSCGQVXCRGEGCVHTASGPHCFP 1369

RESULT 8
 NTG3_RAT
 ID NTG3_RAT STANDARD: PRT: 2319 AA.
 AC 09R172;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Neurogenic locus notch homolog protein 3 precursor (Notch 3).
 GN NOTCH3.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_Taxid=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Haritunians T., Boulter J., Weinmaster G., Schanen N.C.;
 RT "Rattus norvegicus mRNA for Notch 3."
 RN [2]
 RP submitted (SEP-2000) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP FUNCTION.
 RA MEDLINE=21094508; PubMed=11182080;
 RA Tanigaki K., Nogaki F., Takahashi J., Tashiro K., Kurooka H.,
 RA Honjo T.;
 RT "Notch1 and Notch3 instructively restrict bFGF-responsive multipotent
 RT neural progenitor cells to an astroglial fate."
 RN Neuron 29:45-55(2001).
 RN [3]
 RP TISSUE SPECIFICITY.
 RA MEDLINE=21331789; PubMed=11438922;
 RA Iryin D.K., Zurcher S.D., Nguyen T., Weinmaster G., Kornblum H.I.;
 RT "Expression patterns of Notch1, Notch2, and Notch3 suggest multiple
 RT functional roles for the Notch-DSL signaling system during brain
 RT development."
 RN J. Comp. Neurol. 436:167-181(2001).
 RN [4]
 RP FUNCTION: Functions as a receptor for membrane-bound ligands
 CC Jagged1, Jagged2 and Delta1 to regulate cell-fate determination.
 CC Upon ligand activation through the released notch intracellular
 CC domain (NICD) it forms a transcriptional activator complex with
 CC RBP-J kappa and activates genes of the enhancer of split locus.
 CC Affects the implementation of differentiation, proliferation and
 CC apoptotic programs (By similarity). Acts instructively to control
 CC the cell fate determination of CNS multipotent progenitor cells,
 CC resulting in astroglial induction and neuron/oligodendrocyte
 CC suppression.
 CC [5]
 RP SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
 CC terminal fragment N(EC) which are probably linked by disulfide
 CC bonds (By similarity).
 CC [6]
 RP SUBCELLULAR LOCATION: Type I membrane protein. Following
 CC proteolytical processing NICD is translocated to the nucleus.
 CC [7]
 RP TISSUE SPECIFICITY: Expressed in postnatal central nervous system
 CC (CNS) germinal zones and, in early postnatal life, within
 CC numerous cells throughout the CNS. It is more highly localized
 CC to ventricular germinal zones.
 CC [8]
 RP PTM: Synthesized in the endoplasmic reticulum as an inactive form
 CC which is proteolytically cleaved by a furin-like convertase in the
 CC trans-Golgi network before it reaches the plasma membrane to yield
 CC an active, ligand-accessible form. Cleavage results in a C-
 CC terminal fragment N(TM) and a N-terminal fragment N(EC). Following
 CC ligand binding, it is cleaved by TNF-alpha converting enzyme
 CC (TACE) to yield a membrane-associated intermediate fragment called
 CC notch extracellular truncation (NEX). This fragment is then
 CC cleaved by presenilin dependent gamma-secretase to release a
 CC notch-derived peptide containing the intracellular domain (NICD)
 CC from the membrane (By similarity).
 CC [9]
 RP PTM: Phosphorylated (By similarity).
 CC [10]
 RP SIMILARITY: BELONGS TO THE NOTCH FAMILY.
 CC [11]
 RP SIMILARITY: CONTAINS 34 EGF-LIKE DOMAINS.
 CC [12]
 RP SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
 CC [13]
 RP SIMILARITY: CONTAINS 5 ANK REPEATS.
 CC [14]
 RP This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its


```

OY 224 -----LCPGKHGPOCE----- 235
DB 333 DVDECSIGANPCENHIGRCVNTGSGFLCCQGRGTYTPTCTDVNELLSPGCRNQTCLDRI 452
OY 236 -----QRCPCQNGVC--HHVTG--ECSGPGMMGTVC----- 264
DB 453 GQFTICMAGFTGTCFCEVDIDECQSSPCVNGVCNDRVNGFCCTCPSGSSGTCQLDVE 512
OY 265 -----GCP-----CPRGRKNCQSQ---ECQ---CHNGTCTDA--ATGQCHC 298
DB 513 CASTPCRNAGKCVDPDPDYEYECRCAGFEETLCERNVNDSCPPRCHRGRCVDPDIASFSCAC 572
OY 299 SPGTGGERCODE-----CPVGYGYLC-----AETCQ 325
DB 573 APTGTGIRCESQVDECRSPCRRTGCKLDLYDKYLCRCPTGTGVNCEYINDDCASNPT 632
OY 336 ---CVNGKRCYHVSAGACLCCEAGFAGERCEARL-----CPEGLYGIKCDKRCP 369
DB 633 FGVCHDGINRYD-----CVCQPGFTGPLCNVINECCASSPGSGGSGCVDENGFFHC--LCP 686
OY 370 -----CHLENTHS--CHPMG--ECACKRGWGLYCNE----- 398
DB 687 PGLSLPPLCLIPANHPAHRKPSHGVCCHDAPGFGQCVCDPWSGPRCSQSLAPDACESOPQ 746
OY 399 -----TCSPTGYEACQOI--CS---CONGADDSVTGK---CTCAPGFKG 436
DB 747 AGGTCTSDGIGHCACAFEGFQGHCEVLSPTPSLCERHGHCESDPDQLTVCSCPPMGOG 806
OY 437 IDC-----STPC-PLGT--YGINCSRCGCKNDAY-----CSP----- 466
DB 807 PRCQDQDVDECAAGSPGPHGTCTNLPSPRCICHGTYTGFCDQDIDDDPAPCLNGSGC 866
OY 467 VGC-----SCTCAAGHGVDC-----STRCSGTMWGACN-- 496
DB 867 QGVGVSFSCSLSGFAGPCARDVDECLSSPCGPGCTDHDVASFTCTCPGYGFRHCETD 926
OY 497 -LTC---OCLNGACNTLDG---TCTCAPGWRGKC----- 525
DB 927 LLDSPSSCFNGCTC--VDGVNSFSCICRPGYTGTHCYKYVDPCTSRPLHGCICNPTHS 984
OY 526 --ELPCQDGTGYLNCERCD-----CSHADGCHPTGHCRLPWSGVHCD----- 569
DB 985 GFECTCRGFTGNOCNPVDMCSQAPCQNGRCVOTGATCICPPEMSGPLCIDIPSLPCTE 1044
OY 570 -----SYCAGRWGPNCSL---PC 585
DB 1045 AAHMGVRLQLCQAGGOCIDKDHSHYCVCPGRMGSHCEDEVDPC 1090

RESULT 9
TEXT_HUMAN STANDARD: PRT: 4289 AA.
AC P22105; P78530; P78531; Q08424; Q90M67;
DT 01-AUG-1991 (Rel. 19, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tenascin-X precursor (TN-X) (hexabrachion-like).
OS TNXB OR TNX OR XB OR HMBL.
OC Homo sapiens (Human).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., Baskin D., Faust J., Loretz C., Ahearn M.E.,
RA Rowen L., Schwartz S., Smith T.M., Spies T., Hood L.;
RT "Sequence determination of 300 kilobases of the human class III MHC
RT locus."
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-747 AND 1687-1944 FROM N.A.
RC TISSUE=Leukocyte;
DR MEDLINE=93300909; PubMed=7686164;

```

```

RA Bristow J., Tee M.K., Gitelman S.E., Mellon S.H., Miller W.L.;
RT "Tenascin-X: a novel extracellular matrix protein encoded by the human
RT XB gene overlapping P450C21B."
RL J. Cell Biol. 122:265-278(1993).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM XB-SHORT).
RC TISSUE=Adrenal gland;
RA MEDLINE=96015044; PubMed=8530023;
RX Tee M.K., Thomson A.A., Bristow J., Miller W.L.;
RT "Sequences promoting the transcription of the human XA gene
RT overlapping P450C21A correctly predict the presence of a novel,
RT adrenal-specific, truncated form of tenascin-X."
RL Genomics 28:171-178(1995).
RN [4]
RP SEQUENCE OF 1-23 FROM N.A.
RC TISSUE=Fetal adrenal gland;
RX MEDLINE=97081760; PubMed=8923003;
RA Speck M., Barry F., Miller W.L.;
RT "Alternate promoters and alternate splicing of human tenascin-X, a
RT gene with 5' and 3' ends buried in other genes."
RL Hum. Mol. Genet. 5:1749-1758(1996).
RN [5]
RP SEQUENCE OF 3470-4289 FROM N.A.
RX MEDLINE=89367293; PubMed=2475872;
RA Morel Y., Bristow J., Gitelman S.E., Miller W.L.;
RT "Transcript encoded on the opposite strand of the human steroid 21-
RT hydroxylase/complement component C4 gene locus."
RL Proc. Natl. Acad. Sci. U.S.A. 86:6582-6586(1989).
RN [6]
RP DISEASE.
RA Published-11642233;
RX Schalkwijk J., Zweers M.C., Steijnen P.M., Dean W.B., Taylor G.,
RA van Vlijmen I.M., van Haren B., Miller W.L., Bristow J.;
RT "A recessive form of the Ehlers-Danlos syndrome caused by tenascin-X
RT deficiency."
RL New Engl. J. Med. 345:1167-1175(2001).
CC -1- FUNCTION: APPEARS TO MEDIATE INTERACTIONS BETWEEN CELLS AND THE
CC EXTRACELLULAR MATRIX. SUBSTRATE-ADHESION MOLECULE THAT APPEARS TO
CC INHIBIT CELL MIGRATION. MAY PLAY A ROLE IN SUPPORTING THE GROWTH
CC OF EPITHELIAL TUMORS.
CC -1- SUBCELLULAR LOCATION: Extracellular matrix.
CC -1- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; XB (SHOWN HERE) AND
CC XB-SHORT; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN FETAL ADRENAL, IN FETAL
CC TESTIS, FETAL SMOOTH, STRIATED AND CARDIAC MUSCLE. XB-SHORT IS
CC ONLY EXPRESSED IN THE ADRENAL GLAND.
CC -1- DISEASE: ASSOCIATION WITH CONGENITAL ADRENAL HYPERPLASIA.
CC -1- DISEASE: Defects in TNXB are the cause of an autosomal recessive
CC form of Ehlers-Danlos syndrome (EDS) manifesting with
CC hyperextensible skin, hypermobile joints, and tissue fragility as
CC observed in classical EDS, but lacking atrophic scars and delayed
CC wound healing.
CC -1- SIMILARITY: CONTAINS 19 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 32 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 FIBRINOGEN C-TERMINAL DOMAIN.
CC -1- CAUTION: THERE ARE TWO GENES FOR TN-X: TNXA AND TNXB. TNXA IS A
CC PARTIAL GENE WHICH CAN SOMETIMES RECOMBINE WITH TNXB.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U89337; AAB47488.1; -
DR EMBL: AF019413; AAB67981.1; -
DR EMBL: X71923; CA50739.1; -
DR EMBL: Y13782; CA474109.1; -
DR EMBL: Y13783; CA474110.1; -
DR EMBL: U24488; AAB41287.1; -
DR EMBL: U52696; AAC50889.1; -

```



```

QY 505 -----GACNTLDGTCTCAPGWRGKCELP-----CODGTGVLN 537
DB 589 OCPNDCSQHGYCQ--DGVICICMEGYVSDCSIRCPSPNCHGRCBCEGRCLDPEYTGPT 646
QY 538 CAER---CDCSHADGCHPTTGHCRCLPCWGSVHC-----DSV 571
DB 647 CATRNCAPDACCRRGRC--VGVGVLCHVGVGGEDCGQEBPPASACPGCGCPRELACRAGOCV 704
QY 572 CAEGRMGPNCSL 583
DB 705 CVEGRGPDCAI 716

RESULT 10
NTC2_MOUSE
ID NTC2_MOUSE STANDARD: PRT: 2470 AA.
AC 035516; 080941; 006008;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Neurogenic locus notch homolog protein 2 precursor (Notch 2) (Notch
  B)
GN NOTCH2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Thymus;
RA Hamada Y., Higuchi M., Tsujimoto Y.;
RT "Complete amino acid sequence and multiform transcripts encoded by a
  single copy of mouse Notch2 gene.";
RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 316-1518 FROM N.A.
RC STRAIN=C57BL/6 X CBA; TISSUE=Embryo;
RX MEDLINE=93178563; PubMed=8440332;
RA Lardelli M., Lendahl U.;
RT "Notch A and Notch B-two mouse Notch homologues coexpressed in a
  wide variety of tissues.";
RL Exp. Cell Res. 204:364-372(1993).
RN [3]
RP SEQUENCE OF 1765-2153 FROM N.A.
RX MEDLINE=97075110; PubMed=8917536;
RA Milner L.A., Bigas A., Kopan R., Brashem-Stein C., Bernstein I.D.,
  Martin D.I.;
RT "Inhibition of granulocytic differentiation by mNotch1.";
RL Proc. Natl. Acad. Sci. U.S.A. 93:13014-13019(1996).
RN [4]
RP FUNCTION.
RX MEDLINE=99396706; PubMed=10393120;
RA Hamada Y., Kadokawa Y., Okabe M., Ikawa M., Coleman J.R.,
  Tsujimoto Y.;
RT "Mutation in ankyrin repeats of the mouse Notch2 gene induces early
  embryonic lethality.";
RL Development 126:3415-3424(1999).
RN [5]
RP DEVELOPMENTAL STAGE, AND ALTERNATIVE SPLICING.
RX MEDLINE=9533893; PubMed=7609614;
RA Higuchi M., Kiyama H., Hayakawa T., Hamada Y., Tsujimoto Y.;
RT "Differential expression of Notch1 and Notch2 in developing and adult
  mouse brain.";
RL Brain Res. Mol. Brain Res. 29:263-272(1995).
RN [6]
RP POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS OF MET-1699.
RX MEDLINE=21523956; PubMed=11518718;
RA Saxena M.T., Schroeter E.H., Mumm J.S., Kopan R.;
RT "Murine notch homologs (NL-4) undergo presenilin-dependent
  proteolysis.";
RL J. Biol. Chem. 276:40268-40273(2001).
RN [7]
RP POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS OF MET-1699.

```

```

RX MEDLINE=21374376; PubMed=11459941;
RA Mitani T., Taniguchi Y., Aoki T., Hashimoto N., Honjo T.;
RT "Conservation of the biochemical mechanisms of signal transduction
  among mammalian Notch family members.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9026-9031(2001).
CC -1- FUNCTION: Functions as a receptor for membrane-bound ligands
  Jagged1, Jagged2 and Delta1 to regulate cell-fate determination.
  Upon ligand activation through the released notch intracellular
  domain (NICD) it forms a transcriptional activator complex with
  RBP-J kappa and activates genes of the enhancer of split locus.
  Affects the implementation of differentiation, proliferation and
  apoptotic programs (By similarity). May play an essential role in
  postimplantation development, probably in some aspect of cell
  specification and/or differentiation.
CC -1- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
  terminal fragment N(EC) which are probably linked by disulfide
  bonds.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Following
  proteolytical processing NICD is translocated to the nucleus.
CC -1- ALTERNATIVE PRODUCTS: 2 isoforms: 1 (shown here) and 2; may be
  produced by alternative splicing.
CC -1- TISSUE SPECIFICITY: Expressed in the brain, liver, kidney,
  neuroepithelia, somites, optic vesicles and branchial arches, but
  not heart.
CC -1- DEVELOPMENTAL STAGE: Expressed in the embryonic ventricular zone,
  the postnatal ependymal cells, and the choroid plexus throughout
  embryonic and postnatal development.
CC -1- PTM: Synthesized in the endoplasmic reticulum as an inactive form
  which is proteolytically cleaved by a furin-like convertase in the
  trans-Golgi network before it reaches the plasma membrane to yield
  an active, ligand-accessible form. Cleavage results in a C-
  terminal fragment N(TM) and a N-terminal fragment N(EC). Following
  ligand binding, it is cleaved by TNF-alpha converting enzyme
  (TACE) to yield a membrane-associated intermediate fragment called
  notch extracellular truncation (NEXT). This fragment is then
  cleaved by presenilin dependent gamma-secretase to release a
  notch-derived peptide containing the intracellular domain (NICD)
  from the membrane.
CC -1- PTM: Phosphorylated.
CC -1- SIMILARITY: BELONGS TO THE NOTCH FAMILY.
CC -1- SIMILARITY: CONTAINS 34.5 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 LIN/NOTCH REPEATS.
CC -1- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D32210; BAA22094.1; -
DR EMBL: X68279; CAA48340.1; -
DR EMBL: U31881; AAC52924.1; -
DR HSP: P16109; IFSB.
DR MGD: MGI:97364; Notch2.
DR InterPro: IPR002110; ANK.
DR InterPro: IPR000152; Asx_hydroxy1.
DR InterPro: IPR000561; EGF-1like.
DR InterPro: IPR000742; EGF_2.
DR InterPro: IPR001881; EGF_Ca.
DR InterPro: IPR001438; EGF-II.
DR InterPro: IPR000800; Notch.
DR Pfam: PF00008; EGF; 35.
DR Pfam: PF00023; ank; 6.
DR Pfam: PF00066; notch; 2.
DR PRINTS: PRO0010; EGFBL00.
DR PRINTS: PRO1452; NOTCH.
DR SMART: SM00248; ANK; 4.
DR SMART: SM00179; EGF_CA; 22.
DR SMART: SM00001; EGF_1like; 12.
DR SMART: SM00004; NL; 3.

```

DR PROSITE: PS50088; ANK_REPEAT: 4.
 DR PROSITE: PS50297; ANK_REPEAT: 1.
 DR PROSITE: PS00010; ASX_HYDROXYL: 22.
 DR PROSITE: PS00022; EGF_1: 33.
 DR PROSITE: PS01186; EGF_2: 27.
 DR PROSITE: PS01187; EGF_CA: 22.
 DR Receptor; Transcription regulation; Activator; Differentiation;
 Developmental protein; Repeat; ANK repeat; EGF-like domain;
 Transmembrane; Glycoprotein; Signal; Phosphorylation;
 KW Alternative splicing.
 FT SIGNAL 1 25
 FT CHAIN 26 2470
 FT CHAIN 1666 2470
 FT CHAIN 1697 2470
 FT CHAIN 26 1677
 FT TRANSMEM 1678 1698
 FT DOMAIN 1699 2470
 FT DOMAIN 26 63
 FT DOMAIN 64 102
 FT DOMAIN 105 143
 FT DOMAIN 144 180
 FT DOMAIN 182 219
 FT DOMAIN 221 256
 FT DOMAIN 258 294
 FT DOMAIN 296 334
 FT DOMAIN 336 372
 FT DOMAIN 373 411
 FT DOMAIN 413 452
 FT DOMAIN 454 490
 FT DOMAIN 492 528
 FT DOMAIN 530 566
 FT DOMAIN 568 603
 FT DOMAIN 605 641
 FT DOMAIN 643 678
 FT DOMAIN 680 716
 FT DOMAIN 718 753
 FT DOMAIN 755 791
 FT DOMAIN 829 869
 FT DOMAIN 871 907
 FT DOMAIN 909 945
 FT DOMAIN 947 983
 FT DOMAIN 985 1021
 FT DOMAIN 1023 1059
 FT DOMAIN 1061 1097
 FT DOMAIN 1099 1145
 FT DOMAIN 1147 1183
 FT DOMAIN 1185 1221
 FT DOMAIN 1223 1260
 FT DOMAIN 1262 1300
 FT DOMAIN 1302 1345
 FT DOMAIN 1372 1410
 FT REPEAT 1418 1454
 FT REPEAT 1501 1533
 FT REPEAT 1501 1533
 FT REPEAT 1825 1869
 FT REPEAT 1874 1903
 FT REPEAT 1907 1937
 FT REPEAT 1941 1970
 FT REPEAT 1974 2003
 FT REPEAT 2007 2036
 FT DOMAIN 1645 1648
 FT DOMAIN 1992 1995
 FT DOMAIN 2183 2189
 FT DOMAIN 2183 2189
 FT DOMAIN 2425 2428
 FT DOMAIN 2445 2450
 FT DISULFID 28 41
 FT DISULFID 35 51
 FT DISULFID 53 62
 FT DISULFID 68 79
 Query Match 18.8%; Score 677; DB 1; Length 2470;
 Best Local Similarity 24.7%; Pred. No. 5; Be-36;
 Matches 220; Conservative 78; Mismatches 246; Indels 348; Gaps 56;

QY 3 ISLNSLSPICL-----LLCH-----WIGTASPLNIE--DPNVCSHW-----ES 39
 Db 529 IDIDCSSTPCLNAGKACIDHPNGYECQCATGTTGLDCEINIDNCDPDCCHGGQCDGIDS 588
 QY 40 YSVAYQSEYPRHP--DQI--YITS-----CTDIANFKCT-----RHRVSY----- 76
 Db 589 YTCICNPGYMAICSDQIDECYSSPCLNDGRICIDLVNGYQCNCOPGTSGLNCLEINFDCA 648
 QY 77 RTAYRHG--EKTMYRRKSQCCPGF-----YESGEMCV----- 106
 Db 649 SNNPCHGVCVGINRYSCVSPFGTGQCNIDIDECASNPRKRGATCTINDVNGFRCTIPE 708
 QY 107 ----PHC-----ADKCHGRC--IAPTCQCEPMSGTNCSSACDGHWPCHTSR 151
 Db 709 GPHHPSCTSYQNECLSNPCINHGCTGSLGKYKCLDAGVGNCE--YDKN----EGLSN 762
 QY 152 CQCKNAGLNCITGA--CHCAAGFRMGCE--DRC-----EGGTYNDCN--QRCCQ-- 198
 Db 763 -PCQNGGTNNLVNGYRGTCKRGKGYCNQVNDIECASNPLNLTGCTDDVSGYTCMCL 821
 QY 199 --NGATCDHYTEGRCRCPGYTGAFCE-----LCPPGKHGPOCE--QRC--PCQ 241
 Db 822 PYTKNQQTFLAPSPNCEAAYCKEAPNESESCLCAPGMOGRKCTVDVDECTSKPCM 881
 QY 242 NGVYCHVTGE--CSCPSGMNGVYCGPCPRGRFKNCQBC--QCHNGTC--DATG 294
 Db 882 NNGVCHVTGSGYVECEPFGSGMCEEDI-----NCLANPQONGSCVDHVTFF 931
 QY 295 QCHSPYTGRCODE-----CPRTGYVLC--AEHC-- 324
 Db 932 SCQHPGFIGDKCTDNECLSEPCCKNGTCSYVNSTYTCPPAGFGVHCENNIDCTE 991
 QY 325 -QCVNGSKCYHVS--ACICEAGFAG-----ERCEAR-----LC 355
 Db 992 SSCFNNGTC--YDINSFSLCPVGTGPFCLHDINCSSRPLCNACTGVDGLTYKIC 1049
 QY 356 PEGYIGKCD-----KRCQCHLENTSCHPMSECAKPGWSGLYNE----- 398
 Db 1050 PLGYTGKNCQTLVNLCSRSPCKNKGTCVQEKARPHCLCPROMDAYCDVLVNSCKAALQ 1109
 QY 399 -----TCSPGFGEACQO--ICS--CONGADCSVYG-- 426
 Db 1110 KGVYVEHLCHQSHGICINAGNTHHCQCPVGTGTSCEQGLDEBCASNPOCHGTCTDFTIGY 1169
 QY 427 KCTCAPGFKIDCSTPCPLGTGYNCSKCGCKNDAYCSPVDSCTCKAGHGVDCSIRC 486
 Db 1170 RCCEVPGYQGVNCE-----YEVDCEQNPCCQNGGTCIDLVNHFKS-----C 1211
 QY 487 PSGTWGFGC--NL--TC-----QCLNGGAC--NTLDG--TCTCAPGMRGEC-----LPC 529
 Db 1212 PPGTRGLICEENIDECAGPHCLNGGQCVDRIGGYTCRCLPFAGECEEDINECLSNPC 1271
 QY 530 -ODGTGVLNCAE-----RCDCSHA-----DGC-----HPTTGHG 557
 Db 1272 SSBGS--LDCVQLKNNYNCICRSAPFTGRHCEFTFLDVCPPKCLNGTCAVAASNNPQGFIC 1329
 QY 558 RCLPGMSGVHCDVCAAGRW-----GPNC-----SLPC 585
 Db 1330 RCPGFRSGARLQSSCGQVYKCRREGQCIHTDSGRFCLNPKDDESGCASNPC 1381
 RESULT 11
 NOTC_DROME STANDARD; PRT: 2703 AA.
 AC P07207: P04154: 097458: 09M4T8;
 DT 01-NOV-1986 (Rel. 03, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Neurogenic locus Notch protein precursor
 GN N OR EG:140G11.1 OR EG:163A10.2 OR CG3936.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;

RX MEDLINE:99180765; PubMed:10079256;
 RA Gray G.E., Mann R.S., Mitsiadis E., Henrique D., Carcangiu M.-L.,
 RA Banks A., Leiman J., Ward D., Ish-Horowitz D., Aravanis-Tsakonas S.,
 RL "Human ligands of the Notch receptor.";
 RL Am. J. Pathol. 154:785-794(1999).
 CC -1- FUNCTION: Functions as a receptor for membrane-bound ligands
 CC Jagged1, Jagged2 and Delta1 to regulate cell-fate determination.
 CC Upon ligand activation through the released notch intracellular
 CC domain (NICD) it forms a transcriptional activator complex with
 CC RBP-J kappa and activates genes of the enhancer of split locus.
 CC Affects the implementation of differentiation, proliferation and
 CC apoptotic programs. May regulate branching morphogenesis in the
 CC developing vascular system (By similarity).
 CC -1- SUBUNIT: Heterodimer of a C-terminal fragment N(TW) and a N-
 CC terminal fragment N(EC) which are probably linked by disulfide
 CC bonds (By similarity).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Following
 CC proteolytical processing NICD is translocated to the nucleus.
 CC -1- ALTERNATIVE PRODUCTS: 3 isoforms; 1 (shown here), 2 and 3; may be
 CC produced by alternative splicing.
 CC -1- TISSUE SPECIFICITY: Highly expressed in the heart, moderately in
 CC the lung and placenta and at low levels in the liver, skeletal
 CC muscle, kidney, pancreas, spleen, lymph node, thymus, bone marrow
 CC and fetal liver. No expression was seen in adult brain or
 CC peripheral blood leukocytes.
 CC -1- PTM: Synthesized in the endoplasmic reticulum as an inactive form
 CC which is proteolytically cleaved by a furin-like convertase in the
 CC trans-Golgi network before it reaches the plasma membrane to yield
 CC an active, ligand-accessible form. Cleavage results in a C-
 CC terminal fragment N(TW) and a N-terminal fragment N(EC). Following
 CC ligand binding, it is cleaved by TNF-alpha converting enzyme
 CC (TACE) to yield a membrane-associated intermediate fragment called
 CC notch extracellular truncation (NEXT). This fragment is then
 CC cleaved by presenilin dependent gamma-secretase to release a
 CC notch-derived peptide containing the intracellular domain (NICD)
 CC from the membrane (By similarity).
 CC -1- PTM: Phosphorylated (By similarity).
 CC -1- POLYMORPHISM: The poly-Leu region of NOTCH4 (in the signal
 CC peptide) is polymorphic and the number of Leu varies in the
 CC population (from 6 to 12).
 CC -1- SIMILARITY: BELONGS TO THE NOTCH FAMILY.
 CC -1- SIMILARITY: CONTAINS 28 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
 CC -1- SIMILARITY: CONTAINS 5 ANK REPEATS.
 CC -1- CAUTION: Ref.1 sequence differs from that shown due to frameshifts
 CC in position 1438 to 1463.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: D63395; BAA09708.1; ALT_FRAME.
 DR EMBL: D86566; BA13116.1; -
 DR EMBL: U95299; AAC32288.1; -
 DR EMBL: U93335; AAC63097.1; -
 DR EMBL: AB023961; BAB20317.1; -
 DR EMBL: AB024520; BAA88951.1; -
 DR EMBL: AB024578; BAA88952.1; -
 DR HSSP: P08709; 1BP9
 DR Genew: HGNC:7884; NOTCH4.
 DR MIM: 164951; -
 DR InterPro: IPR002110; ANK.
 DR InterPro: IPR000152; Asx_hydroxyl.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR000742; EGF_2.
 DR InterPro: IPR001861; EGF_Ca.
 DR InterPro: IPR001438; EGF_II.
 DR InterPro: IPR000800; Notch.
 DR Pfam: PF00008; EGF; 26.

DR Pfam: PF00023; ank; 6.
 DR Pfam: PF00066; notch; 2.
 DR PRINTS: PR00010; EGFBLD.
 DR PRINTS: PR00011; EGFAMININ.
 DR PRINTS: PR00012; FNTYPEI.
 DR SMART: SM00248; ANK; 5.
 DR SMART: SM00179; EGF_Ca; 11.
 DR SMART: SM00001; EGF_Like; 15.
 DR SMART: SM00004; NL; 2.
 DR PROSITE: PS50088; ANK_REPEAT; 5.
 DR PROSITE: PS50297; ANK_REPEAT_REGION; 1.
 DR PROSITE: PS00010; ASX_HYDROXYL; 11.
 DR PROSITE: PS00022; EGF_1; 28.
 DR PROSITE: PS01186; EGF_2; 21.
 DR PROSITE: PS01187; EGF_Ca; 9.
 KW Receptor; Transcription regulation; Activator; Differentiation;
 KW Developmental protein; Repeat; EGF-like domain;
 KW Transmembrane; Glycoprotein; Signal; Phosphorylation; Polymorphism;
 KW Triplet repeat expansion; Alternative splicing.
 FT SIGNAL 1 23
 FT CHAIN 24 2003
 FT CHAIN 1432 2003
 FT CHAIN 1467 2003
 FT DOMAIN 24 1447
 FT TRANSMEM 1448 1468
 FT DOMAIN 1469 2003
 FT DOMAIN 24 63
 FT DOMAIN 64 115
 FT DOMAIN 118 155
 FT DOMAIN 156 192
 FT DOMAIN 194 232
 FT DOMAIN 234 272
 FT DOMAIN 276 314
 FT DOMAIN 314 353
 FT DOMAIN 355 391
 FT DOMAIN 392 430
 FT DOMAIN 432 473
 FT DOMAIN 475 511
 FT DOMAIN 513 549
 FT DOMAIN 551 587
 FT DOMAIN 589 625
 FT DOMAIN 626 659
 FT DOMAIN 661 689
 FT DOMAIN 691 727
 FT DOMAIN 729 765
 FT DOMAIN 767 803
 FT DOMAIN 806 842
 FT DOMAIN 844 880
 FT DOMAIN 882 928
 FT DOMAIN 930 966
 FT DOMAIN 968 1004
 FT DOMAIN 1006 1044
 FT DOMAIN 1046 1085
 FT DOMAIN 1087 1126
 FT DOMAIN 1130 1171
 FT DOMAIN 1172 1212
 FT REPEAT 1213 1246
 FT REPEAT 1247 1286
 FT REPEAT 1287 1326
 FT REPEAT 1327 1365
 FT REPEAT 1366 1398
 FT REPEAT 1399 1432
 FT REPEAT 1433 1465
 FT REPEAT 1466 1498
 FT REPEAT 1499 1532
 FT REPEAT 1533 1565
 FT REPEAT 1566 1598
 FT REPEAT 1599 1632
 FT REPEAT 1633 1665
 FT REPEAT 1666 1698
 FT REPEAT 1699 1732
 FT REPEAT 1733 1765
 FT REPEAT 1766 1798
 FT DISULFID 28 41
 FT DISULFID 35 51
 FT DISULFID 53 62
 FT DISULFID 68 80
 FT DISULFID 74 103
 FT DISULFID 105 114
 FT DISULFID 122 133
 FT CHAIN 1467 2003
 FT CHAIN 1432 2003
 FT CHAIN 1467 2003
 FT DOMAIN 24 1447
 FT TRANSMEM 1448 1468
 FT DOMAIN 1469 2003
 FT DOMAIN 24 63
 FT DOMAIN 64 115
 FT DOMAIN 118 155
 FT DOMAIN 156 192
 FT DOMAIN 194 232
 FT DOMAIN 234 272
 FT DOMAIN 276 314
 FT DOMAIN 314 353
 FT DOMAIN 355 391
 FT DOMAIN 392 430
 FT DOMAIN 432 473
 FT DOMAIN 475 511
 FT DOMAIN 513 549
 FT DOMAIN 551 587
 FT DOMAIN 589 625
 FT DOMAIN 626 659
 FT DOMAIN 661 689
 FT DOMAIN 691 727
 FT DOMAIN 729 765
 FT DOMAIN 767 803
 FT DOMAIN 806 842
 FT DOMAIN 844 880
 FT DOMAIN 882 928
 FT DOMAIN 930 966
 FT DOMAIN 968 1004
 FT DOMAIN 1006 1044
 FT DOMAIN 1046 1085
 FT DOMAIN 1087 1126
 FT DOMAIN 1130 1171
 FT DOMAIN 1172 1212
 FT REPEAT 1213 1246
 FT REPEAT 1247 1286
 FT REPEAT 1287 1326
 FT REPEAT 1327 1365
 FT REPEAT 1366 1398
 FT REPEAT 1399 1432
 FT REPEAT 1433 1465
 FT REPEAT 1466 1498
 FT REPEAT 1499 1532
 FT REPEAT 1533 1565
 FT REPEAT 1566 1598
 FT REPEAT 1599 1632
 FT REPEAT 1633 1665
 FT REPEAT 1666 1698
 FT REPEAT 1699 1732
 FT REPEAT 1733 1765
 FT REPEAT 1766 1798
 FT DISULFID 28 41
 FT DISULFID 35 51
 FT DISULFID 53 62
 FT DISULFID 68 80
 FT DISULFID 74 103
 FT DISULFID 105 114
 FT DISULFID 122 133


```

FT DISULFID 435 444 BY SIMILARITY.
FT DISULFID 451 461 BY SIMILARITY.
FT DISULFID 455 470 BY SIMILARITY.
FT DISULFID 472 481 BY SIMILARITY.
FT DISULFID 488 499 BY SIMILARITY.
FT DISULFID 493 508 BY SIMILARITY.
FT DISULFID 510 519 BY SIMILARITY.
FT DISULFID 526 537 BY SIMILARITY.
FT DISULFID 531 546 BY SIMILARITY.
FT DISULFID 548 557 BY SIMILARITY.
FT DISULFID 630 641 BY SIMILARITY.
FT DISULFID 635 650 BY SIMILARITY.
FT DISULFID 652 661 BY SIMILARITY.
FT DISULFID 668 679 BY SIMILARITY.
FT DISULFID 673 688 BY SIMILARITY.
FT DISULFID 690 699 BY SIMILARITY.
FT DISULFID 706 717 BY SIMILARITY.
FT DISULFID 711 726 BY SIMILARITY.
FT DISULFID 728 737 BY SIMILARITY.
FT DISULFID 783 794 BY SIMILARITY.
FT DISULFID 788 803 BY SIMILARITY.
FT DISULFID 805 814 BY SIMILARITY.
FT DISULFID 821 832 BY SIMILARITY.
FT DISULFID 826 841 BY SIMILARITY.
FT DISULFID 843 852 BY SIMILARITY.
FT DOMAIN 938 941 POLY-PRO.
FT CARBOHYD 139 139 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 214 214 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 742 742 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 957 957 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 988 988 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1042 1042 N-LINKED (GLCNAC. . .) (POTENTIAL).
SEQUENCE 1213 AA; 133365 MW; 5C5F16A7E20D9534 CRC64;

```

Query Match 18.5%; Score 667.5; DB 1; Length 1213;
 Best Local Similarity 27.7%; Pred. No. 1.4e-35;
 Matches 190; Conservative 64; Mismatches 176; Indels 257; Gaps 45;

```

OY 83 GKTYRRRSQCCPFYEGGEMCVHCHADKCYHRCIAPNTQCPEBGMGTSSACDGD 142
DB 247 GKCYLGMWGPYC-----DKCIPH--PGVHGTCVEPMOCLDITNMGG--OLCDXD 293
OY 143 HMGPHCTSRCCCKNGALCN---PITGACHCAAGFRGMRCE-----DRCEGG 185
DB 294 L--NYCGTHQPLNGTSGNTGPKYQSCEDGYSGVNCERAHACLSPCANGTCKET 351
OY 186 TYGNDCH-----QRC---OCQNGATC-DHYTG-ECRCPPGYTGAFCEDLCP 226
DB 352 SQGYEHCALIGWSTSCSEINVDCTPNQCKHGTGQDLVNGFKCACPRHMTGKTCQ---- 407
OY 227 PEKHPRCEQRRCQNGGCVHVTGE--CSCPSGMGTYCGPCEBGRGKNC---QRC 281
DB 408 --IDANECDK-PCVYANASCHNLIGAYFCECLPGWS-----GQNCDDININDC 451
OY 282 --QCHNGGTC-DAATG-QCHSPGYTGERCQ---DECPYGYGLVLAETCCQVNGKCY- 333
DB 452 KGQCLNGGCKDLVNGYKCLCPRGYTGCEKDVDC-----ASSPLNGRCD 501
OY 334 HVSQ-ACICEAGFAGEKCEARL-----CPBG--LYGICKDKRCP-----HLE 373
DB 502 EYNGEQCLCPAFSGQLQDLIDYCKPNCQAGACFNLASDYFCFCPDYEGKNCSHLK 561
OY 374 N---THSCHPMSC-----EACRKGWGL 394
DB 562 DHCRITSCQVDSCTVAANASTPEGVRISSNVCGPHGRCSRQAGQFTCCQGGFRGT 621
OY 395 YCNETCSPEFYGEACQOQICSCQNGADC-DSV-TGKCTCAPGFKGIDCSTPCLGTGYNIC 452
DB 622 YCHEININD-----CES-NPCRNNGTCLDKVNVYQICADGMEGVHCE-----INDD 667
OY 453 SSRGCKNDAYCSPV--DGSCTCKAGMHGVDCSIF--CPSGTWGFGNLTCCQCLNGAC 507

```

```

DB 668 CSLNPLNKAGACODLVNDFYCEBGRNGKGTCHSDSODEAT-----CNNGGTC 717
OY 508 NTLDFG--CTCAPRGENCE-----LP-----CODGTGLMCA 539
DB 718 HDEGTFKRCRSPGMEGATCNIAKNSCLPNCENGTCVWNGDSFNCYCKEGMEGSTCT 777
OY 540 EHC-DCSHADGCHP--TTG-----HCRCLPGMSGVHC-----568
DB 778 ENTNDONP-----HPCYNSGTCTVDGENMYRCBAPGAPGDCNININEOSSPCAFGSGTCV 833
OY 569 DSV-----CAEGRWGPC-----SLPC 585
DB 834 DEINGYRCICPPRGIRGPDCCQEVVGRPC 860

RESULT 15
NTC2_HUMAN
ID NTC2_HUMAN STANDARD; PRT; 2471 AA.
AC 004721; Q9H240; Q99734;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Neogenic locus notch homolog protein 2 precursor (Notch 2) (hN2).
GN NTC2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Brain;
RA Blumweller C.M., Mann R.S.;
RT Submitted (NOV-2000) to the EMBL/Genbank/DBJ databases.
RL [2]
RN SEQUENCE FROM N.A.
RP TISSUE-Breast tumor;
RC Correia R.G., Camargo A.A., Moreira E.S., Simpson A.J.;
RT "Human Notch2, a novel member of cell-fate determining NOTCH
family";
RL Submitted (OCT-2000) to the EMBL/Genbank/DBJ databases.
RN [3]
RN SEQUENCE OF 967-1229 FROM N.A.
RP TISSUE-T-cell;
RA Lemasson I., Devaux C., Mesnard J.M.;
RT "Partial sequence of EGF-like repeat domain of human Notch2 mRNA";
RL Submitted (NOV-1996) to the EMBL/Genbank/DBJ databases.
RN [4]
RN SEQUENCE OF 1810-2447 FROM N.A.
RP TISSUE-Brain;
RC MEDLINE=93265135; PubMed=1303260;
RA Stifani S., Blumweller C.M., Redhead N.J., Hall R.E.,
RT "Human homologs of a Drosophila enhancer of split gene product define
a novel family of nuclear proteins.";
RN Nat. Genet. 2:119-127(1992).
RN [5]
RN POST-TRANSLATIONAL PROCESSING
RP MEDLINE=97386453; PubMed=9244302;
RA Blumweller C.M., Qi H., Zagouras P., Artavanis-Tsakonas S.;
RT "Intracellular cleavage of Notch leads to a heterodimeric receptor on
the plasma membrane.";
RL Cell 90:281-291(1997).
RN [6]
RN IDENTIFICATION OF LIGANDS.
RP MEDLINE=99180765; PubMed=10079256;
RA Gary G.E., Mann R.S., Mitsiadis E., Henrique D., Carcangiu M.-L.,
RA Barks A., Leiman J., Ward D., Ish-Horowitz D., Artavanis-Tsakonas S.;
RT "Human ligands of the Notch receptor.";
RL Am. J. Pathol. 154:785-794(1999).
CC -!- FUNCTION: Functions as a receptor for membrane-bound ligands
CC Jagged1, Jagged2 and Delta1 to regulate cell-fate determination.
CC upon ligand activation through the released notch intracellular

```


Best Local Similarity 23.3%; Pred. No. 2.7e-35;
Matches 209; Conservative 64; Mismatches 191; Indels 433; Gaps 46;

```
QY 93 QCCPGFYESGEMCVPHCADKCVHGR-CIAPNTCQ-----CEPGMGITNCSSACDG 141
DB 91 RCASGF--TGEDCQYSTSHPCFVSRPLNGTCHMLSRDYTECTQVGTGKRC-----142
QY 142 DHKGPCTSRCCCKNCAICNPITG--ACHCAGFRGRCEDRCQGTGNDCHQRCOCN 199
DB 143 -QMTDACLTH-PCANGSTCTTVAAGPSCKLTGFTGQKE-----TDVNECDIPGHCOH 194
QY 200 GATCDHVTG--ECRCPPGTGAFCEDL-----CPRGKKG 231
DB 195 GGTCLNLPSYQCQCPQGTGQYCDLHYPCAPSPCVNGTCTQGTDTFECNCLPFEBS 254
QY 232 POCQO-----RCP-----CONGV 245
DB 255 STCERNIDCPRNHRCONGVCVDVNTYNCRCPPQTGQFCTEDVDECLLQPNACONGT 314
QY 246 CHHVTG--ECSCPSGMGTVCQO-----PCPEGRFGKNC- 277
DB 315 CANRNGYGCVCYNGWSGDCSENIDCAFASCTPGSTCIDRYASFSCMCPEKAGLICH 374
QY 278 -SQEC--OCHNGTCD--ATGO--CHCSPGYTGERCO--DECPVGTGYVLCATCOC 326
DB 375 LDDACISNPHCHGALCDTNPINQYICTCPQGTGKADCTEDVDECAM-----ANSNFC 427
QY 327 VNGKCYHVSAG--CICERGFAGERCE-----ARLCPEGLY 360
DB 428 EHGKCVNTIDGAFHCECLGAGPCEMDINECHSDPCQNDATCLDKIGFTCLCMPFK 487
QY 361 GIKCDKR-----CP-----CHLE-----NTHSC 378
DB 488 GYHCELEINECOSNFCVNNQCVDKYNRPQCLCPGFTGPVCQIDIDCSSITPCLNGAKC 547
QY 379 --HPMSGECACRPGMSGLYCNET-----CSPGFYGEAC-OQ 411
DB 548 IDHPNGYECQCATGFTGVLCEENIDNCDDPCHHGQCQDIDSYTICINPGYMGALCSDQ 607
QY 412 I-----CSCQNGA-----DCDS--VTG-----K 427
DB 608 IDECYSSPCLNDRCIDLNVGYQCNCQPGTSGVNCENINFDCASNPCIHGICMDGINRYS 667
QY 428 CTCAPGFKGIDC-----SPPCPLGTGGINCSS--RCGCK-----NDAYCS 465
DB 668 CYCSPGFTGQRCNIDIDECASNPCKRKGATCINGVNGFRCICPEGPHHPSCYSOVNECLSN 727
QY 466 P-VDGSC-----CRAGMHGVDCSI-----RCPSTG 490
DB 728 PCIHGNCCTGGLSGYKCLDAGWVINGCEVDKNECLSNPONGGTCDNLVNGYRCTCKKG 787
QY 491 WGFGCMLTQO-----CLNGGAC-----507
DB 788 KGYNCQVINIDECASNPLNQGTCTCFDDISGYTCHVLYTGKNCQTVLAPCSPNPECNAAV 847
QY 508 -----NTLDGTCTCAPMGKEKCEL-----PCOD-----GTY-----GLN 537
DB 848 CKESPMEFESTICLAPGMOGQRTIDIDECISKPCNHHGLCHANTQGSTYMECPPGSGMD 907
QY 538 CAERCD-----CSHADGCHP--TTGHCRLPQMSGVHCDSYCAEGRWGPNC--SLPC 585
DB 908 CEEDIDDCLANPQONGSCMDGVNTPFSCLCPGFTGDKCQTDME-----CLSEPC 958
```

Search completed: May 9, 2003, 11:55:47
Job time : 26.5064 secs

THIS PAGE BLANK (USPTO)